目　　录

第一章　随机事件和概率

§1.1　随机事件的直观意义及其运算

一、必然现象与随机现象

在自然界里，在生产实践和科学实验中，人们观察到的现象大体可归结为两种类型．一类是可事前预言的，即在准确地重复某些条件下，它的结果总是肯定的，或是根据它过去的状态，在相同条件下完全可以预言将来的发展．我们把这一类型现象称之为确定性现象或必然现象．例如重物在高处总是垂直落到地面；在一个大气压下，水在 100°C 时会沸腾；水稻的生长从播种到收割，总是经过发芽、育秧、长叶、吐穗、扬花、结实这几个阶段；在射击时（假设空气阻力等可以忽略）弹道完全由射击的初始条件（如炮弹的初速、发射角和弹道参数等）决定．早期的科学就是研究这一类现象的规律性，所应用的数学工具如数学分析、几何、代数、微分方程等是大家所熟悉的．但人们逐渐还发现另一类型的现象，它是事前不可预言的，即在相同条件下重复进行试验，每次结果未必相同；或是知道它过去的状况，在相同条件下，未来的发展事前却不能完全肯定．这一类型的现象我们称之为偶然性现象或随机现象．如抛掷一个质地均匀的对称的硬币，结果可能是正面向上，或背面向上；新生的婴儿可能是男或是女；在相同海况与气象条件下，某定点海面的浪高时起时伏；当空气阻力等不能忽略时，弹道不能根据初始条件完全确定，可能向不同方向作程度不同的偏移，事前不能肯定．类似的例子还可以举出许多来．

初时人们把这种现象称为"偶然现象"是指它是"不正常的"、"出乎意料的"或者是"原因不明的"，甚至对于迅雷、疾风、陨石、地震等认为是天降的灾难.

是不是这些偶然现象都没有什么规律性可寻呢？事实上并非如此. 人们通过长期的反复观察和实践，逐渐发现所谓不可预言，只是对一次或少数几次观察或实践而言，当在相同条件下进行大量观察时，偶然现象都呈现某种规律，因而也是可以预言的. 例如根据各个国家各个时期的人口统计资料，新生婴儿中男婴和女婴的比例大约总是 1:1. 据传，在我国古代很早的时候就已经知道了这样一个结果. 又如人的高度虽然各不相同，但通过大量的统计，如果在一定范围内把人的高度按所占的比例画出"直方图"，就可以连成一条和铜钟的纵剖面一样的曲线；定点海面在一段时间内的浪高，也可以画出类似的曲线，如图 1.1.1. 还有更简单的例子(大家可以立即检验的)，均匀的硬币抛掷多次，正面和背面出现的次数比例总是近似 1:1，而且大体上抛掷次数愈多，愈接近这个比值. 历史上，蒲丰掷过 4040 次，得到 2048 次正面；皮尔逊掷过 24000 次，得到 12012 次正面.

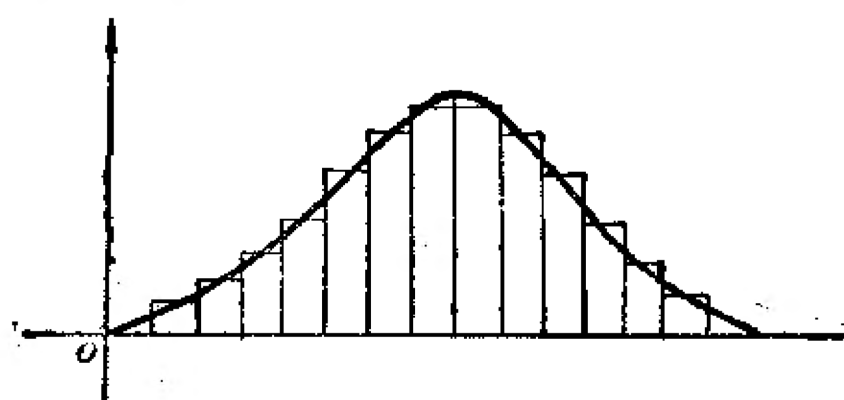


图 1.1.1

从上面的叙述中我们看到，自然界中存在着具有如下特性的现象：在一定的条件组实现时，有多种可能的结果发生，事前人们

不能预言将出现那种结果，但大量重复观察时，所得的结果却呈现某种规律，称为随机现象的统计规律性.

正如革命导师恩格斯所说："在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现这些规律。"(《马克思恩格斯选集》中译本第四卷 243页).

概率论与数理统计就是研究随机现象统计规律性的一门数学学科.

根据马克思、恩格斯的论述，必然性与偶然性是对立统一的概念，偶然性不能理解为"碰巧的"，(为了避免作这种误解，人们往往把它称为"随机性")它蕴含内在必然性的规律；反过来，被断定为必然的东西，是由纯粹的偶然性构成的.

例 1.1.1 研究气体的性质知道，由于气体是由数目众多的分子构成，这些分子以很快的速度进行剧烈的运动，在运动的过程中相互碰撞而改变其动量和方向，因而每个分子的运动是随机现象. 而大量的分子运动呈现出的总体现象——温度、压强却满足波义耳定律.

曾经有人认为，所以会出现事前不可预言的偶然现象，是因为我们对一个现象出现的原因还缺乏全面足够的知识，认为随着科学的发展和人类认识的深化，总有一天将不再存在不可预言的随机现象.

诚然，增加条件组的条件来减少随机性是可能的. 例如，人的高度按不同的年龄、地区、性别是有差异的，若限定在同一地区，在一定年龄范围内选取同性别的人来测量高度，可能会得到比一般任意选取时有较均匀的结果，但随机因素的影响总是不可避免的. 很多现象初始条件的稍微改变，其产生的后果却差别很大. 在实际中即使人们有可能把条件组的条件绝对控制到每次都一样，但影响的因素还是大量的，且互相作用错综复杂. 例如在例 1.1.1 关

于气体分子运动的例子中，即使我们把一立方厘米气体的 2.683×10^{18} 个分子的运动方程和初始条件都列出来，且不说我们要求出这组联立方程的解也是很困难的．更应注意这里在建立方程和确定初始条件时已是忽略了许多次要的，但往往是随机的因素．

因此，偶然性现象是客观存在的，那种否认偶然现象的想法，是"力图用根本否认偶然性的办法来对付偶然性"（恩格斯《自然辩证法》中译本 196 页）．

二、随机试验与事件

为了叙述方便，我们把对某种自然现象作一次观察或进行一次科学试验，统称为一个试验．如果这个试验"在相同条件下可以重复进行"，而且每次试验的结果事前不可预言，我们就称它为一个随机试验．

下面，我们所说的试验都是指随机试验．

进行一个试验总有一个观察的目的，试验中会观察到有多种不同的可能结果．例如抛掷一个质地均匀的硬币，我们的目的是要观察它那一面朝上，这里只有两种不同的结果："正面"或"背面"，至于硬币落在桌面上那一个位置，朝那个方向滚动等不在目的之列，我们不算作结果．

试验的每一个可能结果一般称为随机事件，简称为事件，我们用字母 $A,B,C,\cdots$ 表示．

例 1.1.2 在 $0,1,2,\cdots,9$ 十个数字中任意选取一个，可有十种不同的结果："取得一个数是 0"，…，"取得一个数是 9"．但还有其它可能结果："取得一个是奇数"，"取得一个是大于 4 的数"，"取得一个数是 3 的倍数"等等．

我们把不可能再分的事件称为基本事件．例如在例 1.1.2中，"取得一个数是 0"，"取得一个数是 1"，…，"取得一个数是 9" 都是基本事件．由若干基本事件组合而成的事件称为复合事件．例如

"取得一个数为3的倍数"是一个复合事件，它由"取得一个数是3"，"取得一个数是 6"，"取得一个数是 9"三个基本事件组合而成.

一个事件是否称为基本事件是相对于试验的目的来说的. 例如量度人的身高，一般说，区间$(0,4)$中的任一实数，都可以是一个基本事件，这时，基本事件有无穷个；但如果量度高度是为了了解乘车是否需要买全票、半票或免票，这时就只有三个基本事件了.

三、事件的关系与运算

进行一个试验，有这样或那样的事件发生，它们各有不同的特性，彼此之间又有一定的联系. 下面我们引进事件之间的一些重要关系和运算，这将有利于今后对事件和它的概率的叙述和研究.

1. 在一定条件组下必然发生的事件称为必然事件. 在一定条件组下必然不发生的事件称为不可能事件. 必然事件用符号Ω表示，不可能事件用符号$\varnothing$表示. 把必然事件和不可能事件也算作随机事件，这对我们讨论问题是方便的.

例如，就目前世界上人高来说"人的高度小于 4 米"是必然事件，而"人的高度大于 4 米"则是不可能事件.

2. 如果两个事件A与B不可能同时发生，则称事件A与B为互不相容. 例如必然事件Ω与不可能事件$\varnothing$是互不相容的，又例如在例 1.1.2 中"取得一个数是 0"和"取得一个数是 1"是互不相容的.

如果$A_1, A_2, \cdots, A_n$中的任意两个事件是互不相容的，则称$A_1, A_2, \cdots, A_n$互不相容.

3. 如果事件A发生必然导致事件B发生，则称事件A含于事件B，或称事件B包含事件A，并记为$A\subset B$. 如例 1.1.2 中，令A表示"取得一数为 4 的倍数"，B表示"取得一数为偶然"，则$A\subset B$.

又如对任一事件 A，有 $\varnothing\subset A\subset\Omega$.

$A\subset B$ 的一个等价说法是，如果事件 B 不发生则事件 A 必然不发生.

如果 $A\subset B$ 且 $B\subset A$，则称 A 与 B 相等(或等价)，记为 $A=B$.

4. 如 C 表示"事件 A 与 B 中至少有一个发生"这一事件，称它为 A 与 B 的和(或并)，记为 $C=A\cup B$. 如在例 1.1.2 中，令 A 表示"取出一数为偶数"①，B 表示"取得一数大于 5"，则 $C=A\cup B$ 表示"取得一数或者大于 5，或者是偶数"，即等价于"取出一数为 0，2，4，6，7，8，9 中之一数".

5. 如 D 表示"事件 A 与 B 同时发生"这一事件，称它为 A 与 B 的积(或交)，记为 $D=A\cap B$. 用上例，则 D 表"取得一数为 6 或 8".

由事件积的定义，立即得到：

a) 对任一事件 A，有 $A\cap\Omega=A$, $A\cap\varnothing=\varnothing$;

b) 若 A_1, A_2 互不相容，则 $A_1\cap A_2=\varnothing$.

6. 如 E 表示"事件 A 发生而事件 B 不发生"这一事件，则称 E 为 A 与 B 之差，记为 $E=A-B$. 用前例，$E=A-B$ 表示"取得一个数为 0，2，4 中之一数".

由二事件之差的定义立即得到：对任意事件 A 有

$$A-A=\varnothing;\quad A-\varnothing=A;$$
$$A-\Omega=\varnothing.$$

7. Ω 与 A 之差 $\Omega-A$ 这一事件称为 A 的逆事件，记为 $\bar{A}$，它表示"A 不发生"这一事件. 用前例，$\bar{A}$ 表示"取出一数为奇数".

事件的和与事件的积都可以推广到有限(有穷)多个事件，即 $A=\bigcup_{i=1}^{n}A_i$ 表示"$A_1, A_2, \cdots, A_n$ 中至少有一事件发生"这一事件.

① 为方便起见，以后均把数 0 算作偶数.

$B=\bigcap_{i=1}^{n} B_i$ 表示"$B_1, B_2, \cdots, B_n$ 同时发生"这一事件。

我们有时需要考虑可列无穷多个事件，需要把事件的和与推广到可列无穷多个的情形，即考虑

$$A=\bigcup_{i=1}^{\infty} A_i \text{ 与 } B=\bigcap_{i=1}^{\infty} B_i,$$

记号 $\bigcup_{i=1}^{\infty} A_i$ 表示"可列无穷多个事件 A_i 中至少有一个发生"，而 $\bigcap_{i=1}^{\infty} B_i$ 表示"可列无穷多个事件 B_i 同时发生"。

为了说明这一需要，现举一例：一人进行科学实验，直到试验成功为止，若 A 表实验成功，A_i 表实验到第 i 次才成功，则显然有

$$A=\bigcup_{i=1}^{\infty} A_i.$$

可以验证一般事件的运算满足如下关系：

1）交换律：$A\cup B=B\cup A$，$A\cap B=B\cap A$；

2）结合律：$A\cup(B\cup C)=(A\cup B)\cup C$，
$A\cap(B\cap C)=(A\cap B)\cap C$；

3）分配律：$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$，
$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$；

分配律可以推广到有穷或可列无穷的情形，即

$A\cap(\bigcup_i A_i)=\bigcup_i(A\cap A_i)$，$A\cup(\bigcap_i A_i)=\bigcap_i(A\cup A_i)$；

4）$A-B=A\cap\bar{B}$；

5）对有穷个或可列无穷个 A_i，恒有

$$\overline{\bigcup_i A_i}=\bigcap_i \bar{A}_i,\quad \overline{\bigcap_i A_i}=\bigcup_i \bar{A}_i.$$

四、用集合与几何图形表示事件，样本空间

对于事件及其运算，如果应用点集的概念和几何图示法，则较

直观而易于理解.

联系于每一随机试验的每一基本事件，用一个只包含一个元素ω的单点集$\{\omega\}$表示，由若干基本事件组成的复合事件，则用包含若干个元素的集合表示，由所有基本事件对应的全部元素组成的集合，给它一个特殊的名称，称为样本空间. 由于任一随机试验的结果必然出现全部基本事件之一，这样，样本空间作为一个事件是必然事件. 样本空间仍以Ω表示. 我们称样本空间中的每一个元素为样本点. 这样一来，集合论的知识(参阅附录 I) 就可以用来解释事件和事件的运算.

我们把它们的术语对照列表如下.

表 1.1.1

符　号	集　合　论	概　率　论
Ω	空　间	样本空间；必然事件
$\varnothing$	空　集	不可能事件
$\omega\in\Omega$	Ω中的点(或称元素)	样本点
$\{\omega\}$	单点集	基本事件
$A\subset\Omega$	Ω的子集A	事件A
$A\subset B$	集合A包含在集合B中	事件A含于事件B
$A=B$	集合A与B相等(或等价)	事件A与B相等(或等价)
$A\cup B$	集合A与B之和(或并)	事件A与B至少有一个发生(事件A与B之和或并)
$A\cap B$	集合A与B之交	事件A与B同时发生(事件A与B之积或交)
$\bar{A}$	集合A之余集	事件A的逆事件
$A-B$	集合A与B之差	事件A发生而事件B不发生(事件A与B之差)
$A\cap B=\varnothing$	集合A与B没有公共元素	事件A与B互不相容

如果以平面上的某一矩形表示样本空间，矩形内的每一点表示样本点，则事件的运算可通过平面上的几何图形表示. 如果用两个小圆形表示事件A和B，如图 1.1.2 所示，阴影部分表示事件

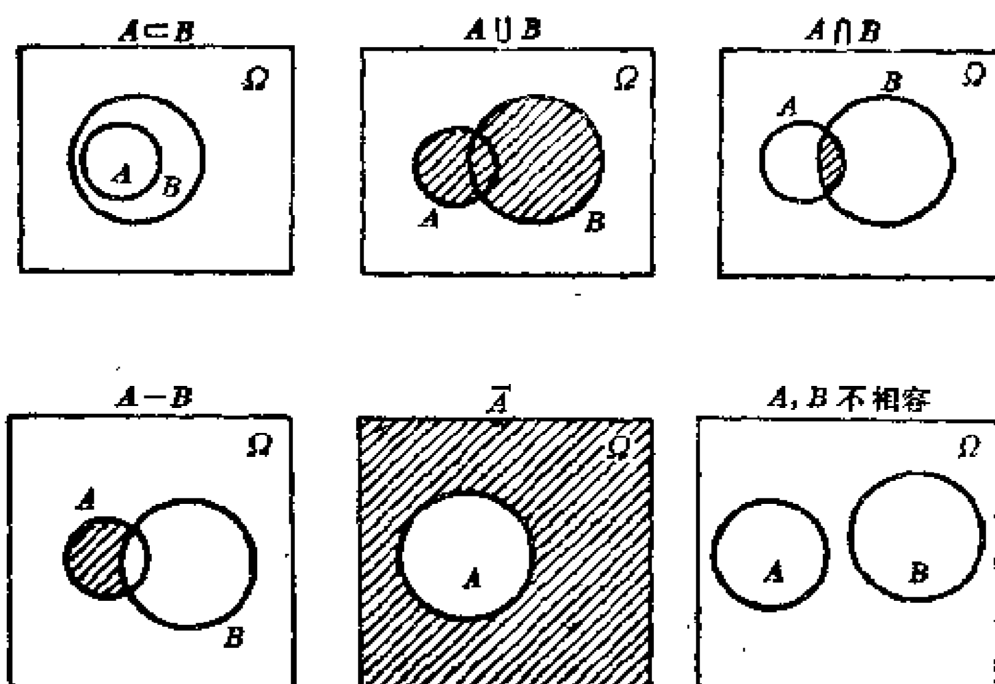


图 1.1.2 ($A\cup B, A\cap B, A-B, \bar{A}$ 分别为图中阴影部分)

A 与 B 的各种关系及运算.

上面说过样本空间是由所有代表基本事件的样本点构成，而基本事件则直接联系于每个随机试验，下面举一些例子说明：

如在例 1.1.2 中样本点是 $\omega_1=0, \omega_2=1, \cdots, \omega_{10}=9$.

基本事件为$\{\omega_i\}$ $(i=1,2,\cdots,10)$；

样本空间$\Omega=\{0,1,\cdots,9\}$.

例 1.1.3 设我们同时抛掷两枚硬币，其基本事件为

$$A_1=\{\omega_1\},\ \ A_2=\{\omega_2\},$$
$$A_3=\{\omega_3\},\ \ A_4=\{\omega_4\},$$

其中 $\omega_1=$(正面，正面)，$\omega_2=$(正面，背面)，$\omega_3=$(背面，正面)，$\omega_4=$(背面，背面)①是样本点. 这时，样本空间由四个样本点构成，即

$$\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}.$$

例 1.1.4 计算某电话站总机在$(0,T]$内的呼叫次数，则基本

① (背面，背面)表示第一个硬币出现背面，第二个硬币也出现背面，余类推.

事件为 $A_0=\{\omega_0\}, A_1=\{\omega_1\}, \cdots, A_n=\{\omega_n\}, \cdots$
其中 $\omega_k=(k$ 次呼叫$)(k=0,1,2,\cdots)$，样本空间为 $\Omega=\{\omega_k: k=0, 1,2,\cdots\}$.

例 1.1.5 设我们测量某一零件时，考虑其测量结果与真正长度的误差. 一般说，它可用一数 x 表示，基本事件为 $\{x\}(-\infty<x<\infty)$①，实数轴上每一点都是其样本点，整个数轴为其样本空间，即 $\Omega=\{x: x\in R_1\}=R_1$.

例 1.1.6 若在例 1.1.5 中同时考虑测量某零件的长度 x 和直径 y，基本事件为 $\{(x,y)\}(0<x<\infty,\ 0<y<\infty)$. 则每一样本点为坐标平面上第一象限的一个点，即数对 $(x,y)(0<x<\infty, 0<y<\infty)$. 这样，其样本空间就是坐标平面的第一象限，即 $\Omega=\{(x,y): 0<x<\infty, 0<y<\infty\}$.

§1.2 概率的直观意义及其计算

我们观察一个随机试验的各种事件，一般来说，总会发现有些事件出现的可能性大些，有些事件出现的可能性小些，有些事件出现的可能性彼此大致相同.

研究随机现象不仅要知道它可能出现那些事件，更重要的是要研究各种事件出现可能性大小，揭示出现这些事件的内在的统计规律，只有这样才有利于我们认识世界和改造世界. 例如知道了某电话总机在 24 小时内出现的呼叫次数的可能性大小，就可以根据要求配置一定的线路设施、管理人员等.

如果只能大概地比较事件发生的可能性大小，要想进行确切的推断，得出精确结论是不行的. 我们要求有一个刻划事件发生可能性大小的数量指标，这个数量指标至少应该满足两个要求:

(i) 它应具有一定的客观性，不能随意改变. 而且理论上可

① 全体实数所成的集记为 R_1.

以通过在"相同条件下"大量的重复试验予以识别和检验.

(ii) 它必须符合一般常情. 例如，事件发生可能性大的，它的值就大;事件发生可能性小的,它的值就小;必然事件的值最大,不可能事件的值最小而等于零.

我们把刻划事件发生可能性大小的数量指标叫做事件的概率. 事件A的概率以$P(A)$表示,并且规定$0\leqslant P(A)\leqslant 1$.

在概率论的发展历史上,人们曾针对不同的问题,从不同的角度给出了定义概率和计算概率的各种方法. 然而所定义的概率都存在一定的缺陷,所以我们毋宁说它是一些计算概率的方法.

一、古典概率

以抛掷质地均匀的硬币为例，人们自然想到由于硬币两面是对称的,所以出现"正面"和"背面"的可能性都一样,人们有理由认为出现"正面"和"背面"的概率都是$\frac{1}{2}$.

在古代较早的时候,在某些特殊情形下,人们利用研究对象的物理或几何性质所具有的对称性,确定计算概率的一种方法如下:对于某一随机试验，如果它的全体基本事件$E_1,E_2,\cdots,E_n$是有穷的,且具有等可能性,则对任意事件A，对应的概率$P(A)$由下式计算:

$$P(A)=\frac{\text{事件}A\text{包含的基本事件数}(k)}{\text{基本事件总数}(n)}, \qquad (1.2.1)$$

并把它称作古典概率.

例1.2.1 设电话号码由0，1,2,…，9共十个数字中任意五个数字组成,设某一户的电话号码是51710,问当不知道这个电话号码时,一次拨号就能拨对该电话号码的概率多少?

解 依题意全部电话号码有10^5个,当不知道电话号码时,拨10^5个电话中的任一电话号码的可能性是相等的. 令A表示"一次

拨号就能拨对该用户号码"这一事件，则这时全部基本事件数 $n=10^5$，而事件 A 只包含一个基本事件，即 $k=1$，按古典概率计算得 $P(A)=\frac{1}{10^5}=0.00001$.

可见当不知道该用户电话号码时，一次拨号就能拨对该电话号码的可能性是很小的.

例 1.2.2 100 只同批生产的外形一样、同型号的三极管中按电流放大系数分类，有 40 只属于甲类，60 只属于乙类. 在按下列两种抽取方法中，求下列事件的概率：

A ="从 100 只中任意抽取 3 只，3 只都是乙类"；

B ="从 100 只中任意抽取 3 只，其中有两只是甲类，1 只是乙类".

抽取方法是

(1) 每次抽取一只，测试后放回，然后再抽取下一只（这种抽取方法称为有返回抽样）；

(2) 每次抽取一只，测试后不放回，在剩下的三极管中再抽取下一只（这种抽取方法称为不返回抽样）.

解 先求事件 A 的概率：

(1) 由于每次抽取测试后均放回，因此，每次都是从 100 个三极管中抽取. 从 100 只中任取 3 只的所有可能取法有 100^3 种，即基本事件总数

$$n=100^3.$$

现考虑事件 A 所包含的基本事件数. 由于属于乙类的三极管有 60 个，因此抽取 3 只都是乙类的所有可能取法有 60^3 种，即 A 包含的基本事件数为

$$k=60^3.$$

按古典概率计算得

$$P(A)=\frac{60^3}{100^3}=0.216.$$

(2) 由于抽取测试后不放回，因此第一次从 100 只中抽取，而第二次抽取只能在剩下的 99 个中抽取，第三次则从 98 个中抽取，故这时的基本事件数

$$n=100\times99\times98=A_{100}^3.$$

同理，可求得 A 包含的基本事件数

$$k=60\times59\times58=A_{60}^3.$$

所以，不返回抽样时

$$P(A)=\frac{A_{60}^3}{A_{100}^3}\approx0.212.$$

下面求事件 B 的概率：

(1) 此时基本事件总数仍为 $n=100^3$，但 B 包含的基本事件数为

$$k=C_3^2\times40^2\times60.$$

故得

$$P(B)=C_3^2\times\frac{40^2\times60}{100^3}=0.288.$$

(2) 此时基本事件总数仍为 $n=100\times99\times98$，而 B 包含的基本事件数为

$$k=C_3^2\times60\times40\times39.$$

故得

$$P(B)=C_3^2\times\frac{60\times40\times39}{100\times99\times98}\approx0.289.$$

一般说，有返回抽样与不返回抽样计算的概率是不同的，特别在抽取的对象数目不大时更是如此．如若例中的三极管只有10只，甲、乙类的比例不变，即甲类有 4 只，乙类有 6 只，可求得在情况 (1) 和 (2)时，事件 A 的概率 $P(A)$ 分别为 0.216 和 0.166．从例

1.2.2 我们看到，当被抽取的对象数目较大时，有返回抽样和不返回抽样所计算的概率相差不大．人们在实际工作中常利用这一点，把抽取对象数量较大时的不返回抽样（例如破坏性试验——发射炮弹；灯泡的寿命试验等），当作有返回抽样来处理，因为后者在一般情况下计算概率较简单．以后将会看到，这样给分析问题和解决问题带来许多方便．

例 1.2.3 十个号码：1 号，2 号，…，10 号，装于一袋中，从其中任取三个，问大小在中间的号码恰为 5 号的概率是多少？

解 十个号码任取三个共有 C_{10}^3 种取法，而三个中，欲大小在中间的号码为 5 号，必须一个小于 5，一个等于 5，一个大于 5，这样的取法有 $C_4^1C_1^1C_5^1$ 种，因此所求的概率为

$$P=\frac{C_4^1C_1^1C_5^1}{C_{10}^3}=\frac{1}{6}.$$

例 1.2.3 的更一般提法为：一袋中有 n 个球，其中 n_1 个带有号码"1"，n_2 个带有号码"2"，…，n_k 个带有号码"k"，$n_1+n_2+\cdots+n_k=n$．从此袋中任取 m 个球，求恰有 m_i 个带有号码"i"$(i=1,2,\cdots,k)$的概率，其中 $m_1+m_2+\cdots+m_k=m$．

读者可用例 1.2.3 同样的解法，得所求概率为

$$P=\frac{C_{n_1}^{m_1}\cdot C_{n_2}^{m_2}\cdot C_{n_3}^{m_3}\cdots C_{n_k}^{m_k}}{C_n^m}.$$

上式称为超几何概率，实际问题中的许多问题都可套用这一模型．

例 1.2.4 麦克斯威尔-波尔茨曼质点运动问题．

设有 m 个不同的质点，每一质点以等可能落于 $N(N\geqslant m)$ 个盒子中的每一个盒子里（设每个盒子能容纳的质点数是没有限制的），求事件 $A=$"某预先指定的 m 个盒子各含一个质点"的概率．

解 由假定，每一质点都以等可能落于 N 个盒中的任一个，故

m 个质点落于 N 个盒子的全部形式有 N^m 种，即全体基本事件数 $n=N^m$，而某预先指定的 m 个盒子中各含一个质点的可能形式有 $m(m-1)(m-2)\cdots 2\cdot 1=m!$，即事件 A 所含的基本事件数 $k=m!$. 依(1.2.1)式得

$$P(A)=\frac{m!}{N^m}.$$

例 1.2.5 设有 r 个人，$r\leqslant 365$. 并设每人的生日在一年365天中的每一天的可能性是均等的. 问此 r 个人有不同生日的概率是多少？

解 r 个人都以等可能的机会在 365 天中的任一天出生，故此题中的基本事件总数为 365^r. 依题意，所求事件所含的基本事件数为

$$365\times 364\times\cdots\times(365-r+1),$$

它恰是 365 个数中任取 r 个数的排列 $A_{365}^r=\frac{365!}{(365-r)!}$.

故所求的概率为

$$P=\frac{365!}{(365-r)!\,365^r}=\frac{365\times 364\times\cdots\times(365-r+1)}{365^r}$$

$$=\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{r-1}{365}\right).$$

当 r 较小时，$P\approx 1-\frac{1}{365}-\frac{2}{365}-\cdots-\frac{r-1}{365}=1-\frac{r(r-1)}{730}$. 如 $r=2$ 时，得"两人不同生日"的概率 $P\approx 1-\frac{2}{730}\approx 0.997$.

从古典概率计算公式(1.2.1)可以得到下面的性质：

性质 1.2.1 对古典概率有

1) 设 A 为任一事件，则 $0\leqslant P(A)\leqslant 1$;

2) 对必然事件 Ω，有 $P(\Omega)=1$;

3) 设事件 $A_1, A_2, \cdots, A_m$ 互不相容，则

$$P\left(\bigcup_{i=1}^{m} A_i\right)=\sum_{i=1}^{m} P(A_i).$$

证 1) 因为任一事件 A 所含的基本事件数 k 总满足

$$0\leqslant k\leqslant n,$$

故有

$$0\leqslant \frac{k}{n}\leqslant 1.$$

2) 由于必然事件由全部 n 个基本事件所组成,即必然事件 Ω 所包含的基本事件数 $k=n$,根据(1. 2. 1)式得

$$P(\Omega)=\frac{n}{n}=1.$$

3) 设 A_i 含有 $k_i(k_i\leqslant n)$ 个基本事件, $i=1, 2, \cdots, m$. 由(1. 2. 1)式得

$$P(A_i)=\frac{k_i}{n} \qquad i=1,2,\cdots,m.$$

由于 A_i 互不相容,则 $\bigcup\limits_{i=1}^{m} A_i$ 含有 $\sum\limits_{i=1}^{m} k_i$ 个不相同的基本事件,由此

$$P\left(\bigcup_{i=1}^{m} A_i\right)=\frac{1}{n}\sum_{i=1}^{m} k_i=\sum_{i=1}^{m}\frac{k_i}{n}=\sum_{i=1}^{m} P(A_i). \blacksquare$$

在古典概率计算过程中,"基本事件是等可能的"是个基本假设,在实际问题中,往往不知道是否满足这个假设,而只凭主观对物理性质或几何的对称性来判断是不完全确切的. 例如一个均匀六面体,分别刻有 1 点至 6 点,假设点的大小相同. 如果认为六个面大致是对称的,因此随意抛掷六面体的结果,出现 1 点,2 点,…, 6 点的可能性似乎都是$\frac{1}{6}$. 但多次抛掷的结果将表明: 刻有 4 点、5 点、6 点这三面出现的总次数比刻有 1 点、2 点、3 点的总次数要多, 这是因为六面体的重心向刻有 1 点 2 点 3 点的一角稍偏移的

缘故.

因此,人们认为要判定事件发生可能性大小,最可靠的办法是通过重复多次试验,特别当基本事件不可能判定是否为等可能性时尤其要采用这个办法,这就提出了统计概率这一名称(也曾称为经验概率).

二、统计概率

在§1.1节中,我们可以看到,在抛掷一个质地均匀的硬币时,出现正面的次数与抛掷总次数之比(我们称它为频率)约为$\frac{1}{2}$,同样新生婴孩中男婴占总数的比(频率)约为$\frac{1}{2}$.我们还可以看下面的例子:

例 1.2.6 检查某工厂产品,其结果如下:

表 1.2.1

抽出产品件数(n)	5	10	60	150	600	900	1200	1800	2400
次品数 (m)	0	3	7	19	52	100	109	169	248
次品率$\left(\frac{m}{n}\right)$	0	0.3	0.117	0.127	0.087	0.111	0.091	0.094	0.103

从表1.2.1中的数字看出,抽到次品数的多少,具有偶然性,但随着抽样的大量进行,抽取的件数逐渐增多,则可发现次品率在0.1附近摆动.

例 1.2.7 考虑某种子的发芽率.从一大批种子中抽取10批种子做发芽试验,其结果如下

表 1.2.2

种子粒数	2	5	10	70	130	310	700	1500	2000	3000
发芽粒数	2	4	9	60	116	282	639	1339	1806	2715
发芽率	1	0.8	0.9	0.857	0.892	0.910	0.913	0.893	0.903	0.905

从表 1.2.2 看出，发芽率在 0.9 附近摆动.

从上面的例子可看到，它们有如下的特点：当我们考虑事件 A 发生的可能性大小时，只要我们在同一条件组下作大量的重复试验，事件 A 发生的频率(事件 A 发生的次数与试验的总次数之比)呈现某种稳定现象．一般说来，当试验的次数增加时，事件 A 发生的频率总是稳定于某一数附近，而偏离的可能性很小．频率具有“稳定性”这一事实，说明了刻划事件 A 发生可能性大小的数——概率具有一定的客观存在性①.

在处理实际问题时，通常我们是用试验次数足够大时的频率来量度概率的．这正如一根棒在固定的温度等条件下具有称之为“长度”的不变的数值，而在实用上则通常用某个适当的测量值作为其长度(例如用多次测量结果的平均数，或是一次测量的结果).

设在同一条件组下进行了 n 次试验(其中 n 足够大)，事件 A 发生了 m 次，则事件 A 发生的频率 $f(A)$ 定义为

$$f(A)=\frac{A\text{ 出现的次数 }m}{\text{试验的总次数 }n}. \tag{1.2.2}$$

我们用 $f(A)$ 作为事件 A 的概率的一个量度，这样计算的概率称为统计概率.

例如：一个射手射击 500 次，中靶 200 次，我们就说他中靶的概率是 $\frac{2}{5}$；新生的婴儿 10000 人中死亡 4 人就说婴儿死亡率(死亡的概率)是万分之四.

我们可以验证，当试验次数 n 固定时，事件 A 发生的频率 $f(A)$ 有如下性质：

性质 1.2.2

1) $0\leqslant f(A)\leqslant 1$;

① 严格的说来，这是一个理想的模型，因为我们在实际上并不能绝对保证在每次试验时，条件组都保持完全一样，这只是一个理想的假设.

2) $f(\Omega)=1, f(\varnothing)=0$;

3) 若事件 $A_1, A_2, \cdots, A_k$ 互不相容，则

$$f\left(\bigcup_{i=1}^{k} A_i\right)=\sum_{i=1}^{k} f(A_i).$$

证 1)及 2)是显然的，现证 3).

只证 $k=2$ 时的情形.

设在 n 次试验中，A_1 发生 m_1 次，A_2 发生了 m_2 次，因而 $f(A_1)=\frac{m_1}{n}, f(A_2)=\frac{m_2}{n}$. 由于 A_1 与 A_2 互不相容，因此事件 $A_1 \cup A_2$ 发生的频率为

$$f(A_1 \cup A_2)=\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}=f(A_1)+f(A_2).$$

因此，统计概率满足

1) $0 \leqslant P(A) \leqslant 1$;

2) $P(\Omega)=1, P(\varnothing)=0$;

3) 若 $A_1, A_2, \cdots, A_k$ 互不相容，则

$$P\left(\bigcup_{i=1}^{k} A_i\right)=\sum_{i=1}^{k} P(A_i).$$

统计概率同样具有理论上和应用上的缺点. 因为我们没有理由认为，取试验次数为 $n+1$ 来计算频率，总会比取试验次数为 n 来计算频率将会更准确、更逼近所求的概率. 在实际应用上，我们不知道 n 取多大才行，如果 n 要很大，我们不一定能保证每次试验的条件都完全一样(如射手射击一千次，很难保证每次射击时的条件都完全一样).

三、几何概率

在概率论发展的早期，就已经注意到只考虑随机现象的可能结果只有有穷个基本事件是不够的，还必须计算有无穷个基本事

件的情形.

设联系于某一随机现象的样本空间，可用欧氏空间的某一区域S表示，其样本点具有所谓"均匀分布"的性质. 这里所说的"均匀分布"类似于古典概率中的等可能性这一概念. 下面关于"均匀分布"暂时只作解释，不作数学上的精确定义. 在我们所述的问题中，总是假设区域S以及其中任一可能出现的小区域A都是可以量度的，其度量大小用$\mu(A)$表示. 例如一维区间的长度，二维区间的面积，三维空间中的体积，…. 并且假定这种度量具有如长度一样的各种性质，如量度的非负性、可加性等.

设某一事件A(也是某一区域)，$A\subset S$，它的度量大小为$\mu(A)$，若以$P(A)$表示事件A发生的概率，考虑到"均匀分布"性，事件A发生的概率取为

$$P(A)=\frac{\mu(A)}{\mu(S)}. \tag{1.2.3}$$

这样计算的概率，称为几何概率.

若$\varnothing$是不可能事件，即$\varnothing$为S中的空的区域，其量度大小为0，故其概率

$$P(\varnothing)=0.$$

例 1.2.8 在时间间隔$[0,T]$内的任何瞬间，两不相关的信号均等可能地进入收音机. 如果当这两个信号进入收音机的间隔时间不大于t，则收音机就受到干扰. 试求收音机受到干扰的概率.

解 以x及y分别表两信号进入收音机的瞬间，由假定，

$$0\leqslant x\leqslant T,$$

$$0\leqslant y\leqslant T.$$

则样本空间是由点(x, y)构成的边长为T的正方形Ω，其面积为T^2，如图 1.2.1 所示. 依题意，收音机受到干扰的充分必要条件为

$$|x-y|\leqslant t.$$

这区域如图 1.2.1 中的 A，这区域位于区域 Ω 内直线 $x-y=t$ 及 $x-y=-t$ 之间，其面积为 $A=T^2-(T-t)^2$. 由(1.2.3)式得所求概率为

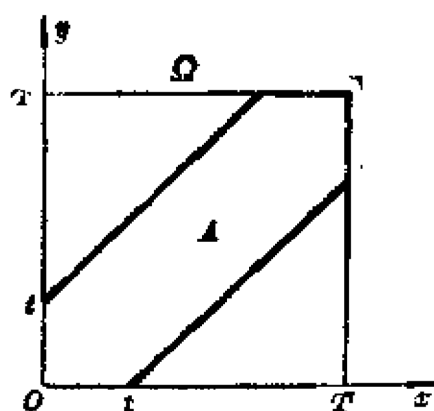


图 1.2.1

$$P=\frac{T^2-(T-t)^2}{T^2}=1-\left(1-\frac{t}{T}\right)^2.$$

从上式可以看到，当 t 相对 T 来说很小时，即假设两信号进入收音机的间隔很短时才产生干扰，而可能进入收音机的时间间隔较长时，收音机受干扰的概率 $P\approx 0$；另一方面，当 $t\approx T$ 时，则有 $P\approx 1$，它表明受干扰的间隔时间 t 与信号可能进入收音机的时间间隔 T 差不多相等时，则收音机以概率为 1 地受干扰. 以上结果在直观上是明显的.

例 1.2.9 (蒲丰的针问题)

在平面上画有等距离为 $a(a>0)$ 的一些平行线，向平面上随意投掷一长为 $l(l<a)$ 的针. 试求针与一平行线相交的概率 P.

解 令

M 表示针的中点；

x 表示针投在平面上时，M 与最近一条平行线的距离；

φ 表示针与最近一条平行线的交角.

如图 1.2.2，容易看出：

$$0\leqslant x\leqslant\frac{a}{2},$$

$$0\leqslant\varphi\leqslant\pi.$$

假定我们取直角坐标图示，则上式表示 $xO\varphi$ 坐标系中的一个矩形 R，如图 1.2.3，而

$$x\leqslant\frac{l}{2}\sin\varphi$$

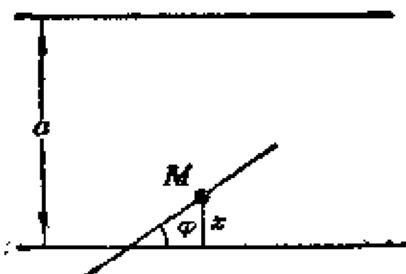


图 1.2.2

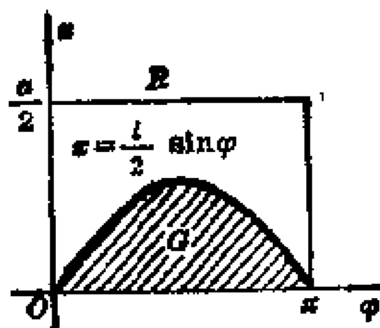


图 1.2.3

是使针与平行线(此线必为与M点最近的平行线)相交的充分必要条件，上面不等式表示图 1.2.3 中阴影部分. 我们把抛掷针到平面上这件事理解为具有"均匀性". 因此，这问题等价于向区域 R 中"均匀分布"地投掷点，而求点落入 G 中的概率 P，由式(1.2.3)得

$$P=\int_0^{\pi}\frac{l}{2}\sin\varphi d\varphi\Big/\frac{a}{2}\pi=2\frac{l}{\pi a}.$$

在计算几何概率时，一开始我们就假定点具有所谓"均匀分布"(类似古典概型中的等可能性)，这一点在求具体问题中的概率时，必须十分注意，否则可能得出不同或者完全错误的结果.

*例 1.2.10 (贝特朗奇论) 在一半径为 r 的圆 C 内"任意"作一弦①，试求此弦长度 l 大于圆内接等边三角形的边长 $\sqrt{3}r$ 的概率 p.

解一 作半径为 $\frac{r}{2}$ 的同心圆 C_1，设弦 AB 的中点 M"任意"落于圆 C 内，如图 1.2.4a. 可以证明，若 M 落于圆 C_1 内，则 $l>\sqrt{3}r$. 由式(1.2.3)，概率 p 为两圆面积之比，即

$$p=\frac{\pi\left(\frac{r}{2}\right)^2}{\pi r^2}=\frac{1}{4}.$$

① 不考虑直径.

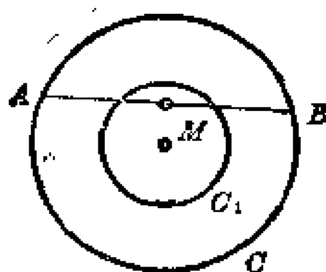


图 1.2.4a

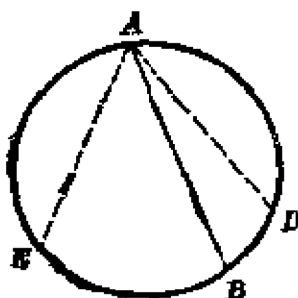


图 1.2.4b

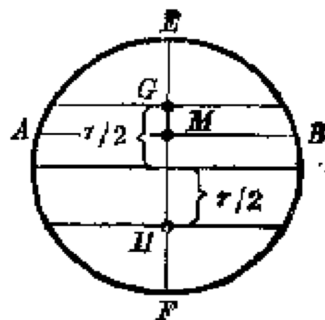


图 1.2.4c

解二 设弦 AB 的一端 A 固定于圆周上，如图 1.2.4 b，另一端 B 是任意的. 考虑等边三角形 $\triangle ADE$，如 B 落于角 A 对应的弧 $\overset{\frown}{DE}$ 上，则 $l>\sqrt{3}r$. 故

$$p=\frac{\overset{\frown}{DE}\text{的弧长}}{\text{圆周长}}=\frac{1}{3}.$$

解三 设弦 AB 垂直于直径 EF，如果 AB 的中点 M 在 GH 上，如图 1.2.4c，则 $l>\sqrt{3}r$，因此

$$p=\frac{GH\text{之长度}}{EF\text{之长度}}=\frac{1}{2}.$$

这样，我们得到了三个不同的答案. 从三种解的过程看到，问题是在于"任意作弦"的提法太不确定了，我们作了上述三种解释，得到了三个不同的结果. 其实，上述三种解法中，所反映的随机事件是不同的. 解一中求得的概率是"随机点 M 落在圆 C_1 中"的概率；解二中求得的概率是"随机点 B 落于圆弧 $\overset{\frown}{DE}$ 上"的概率；而解三则是"随机点 M 落于线段 GH 上"的概率. 因此，相对于每种解释，其计算结果都是正确的.

几何概率有如下性质.

性质 1.2.3 几何概率满足

1) 对任一事件 A 有 $0\leqslant P(A)\leqslant 1$;

2) 对必然事件 Ω 有 $P(\Omega)=1$;

3) 设 $A_1, A_2, \cdots, A_n$ 为互不相容事件，则

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i).$$

证 1)及2)立即可从(1.2.3)式得到. 3)的证明则由度量可加性得到：由于 A_i 的量度为 $\mu(A_i)$，而 A_i 互不相容，故得

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i).$$

但几何概率还具有一个性质（是前面讲古典概率和统计概率时未及考虑的），即完全可加性：

3)′ 设 $A_1, A_2, \cdots, A_n, \cdots$ 为可列无穷多个互不相容事件，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i).$$

例1.2.11 设一质点随机地落于 $I=[0,1]$ 线段内，把 I 分为

$$A_1=\left(\frac{1}{2}, 1\right], A_2=\left(\frac{1}{4}, \frac{1}{2}\right], \cdots A_n=\left(\frac{1}{2^n}, \frac{1}{2^{n-1}}\right], \cdots,$$

规定质点落于这些区间的概率等于线段的长度，即

$$P(A_n)=\frac{1}{2^n} \quad (n=1,2,\cdots).$$

这时

$$I=\bigcup_{i=1}^{\infty} A_i+\{0\}.$$

显然有（注意 $P\{0\}=0$）

$$P(I)=P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i)=\sum_{i=1}^{\infty} \frac{1}{2^i}=1.$$

§1.3 概率模型与公理化结构

上面我们讲了针对不同的问题，分别用古典概率、统计概率和几何概率计算概率的方法，我们举出了它们适用的范围和理论上的缺陷及应用上的局限性. 但这并不能否定他们在具体问题中计

算概率的作用，只是说上述任何一个定义作为概率的数学定义来建立起概率理论是不可能的．但我们看到它们从各自的定义出发都有共同的属性(性质 1.2.1，性质 1.2.2 及性质 1.2.3)，这些从客观事实总结出来的共同属性，可以作为建立概率的数学理论的基础．

概率论也应该象几何、代数一样，通过建立公理化结构，给予概率以数学定义，它既可包括前面三种特殊情况，又具有更广泛的一般性．

前面讲过，联系于一个随机试验首先有一个样本空间 Ω，它是由所有代表基本事件的样本点的全体组成．Ω的子集叫事件．人们总是希望通过用简单的事件的概率来推算复合事件的概率．但是否所有基本事件都能算出它的概率？当基本事件的概率已确定，是否任何Ω的子集的概率都能确定？如果样本空间只有有穷个样本点时(此时称为有穷样本空间)，答案是肯定的，但是对任意样本空间却不一定这样．事实上也不必这样要求，只是对相当广泛的一类事件(Ω的子集或样本点组成的集合)能定义概率就够了．

我们先引进代数与 σ-代数的概念．

定义 1.3.1 设Ω是样本空间，$\mathscr{F}$是由Ω的一些子集所成的集合族，如果$\mathscr{F}$满足如下条件：

i) $\Omega\in\mathscr{F}$；

ii) 若 $A\in\mathscr{F}$，则 $\bar{A}\in\mathscr{F}$；

iii) 若 $A_i\in\mathscr{F}\,(i=1,2,\cdots,n)$，则 $\bigcup\limits_{i=1}^{n} A_i\in\mathscr{F}$，

则称$\mathscr{F}$为体，或叫代数．

定义 1.3.2 如$\mathscr{F}$除了满足上述条件 i)，ii)外，还满足

iii)′ 若 $A_i\in\mathscr{F}\,(i=1,2,\cdots)$，则 $\bigcup\limits_{i=1}^{\infty} A_i\in\mathscr{F}$，

则称$\mathscr{F}$为 σ-代数．$\mathscr{F}$中的集合称为事件．$\mathscr{F}$也称为事件体．

如果样本空间是有穷的，则它的任何代数也必是 σ-代数．

可以验证，对任意样本空间 Ω，它的所有子集（包括空集$\varnothing$及 Ω）所成的集合族就是一个 σ-代数．

例 1.3.1 在编号为 $1,2,\cdots,n$ 的 n 部电视机中任选一部．

（1）若考虑电视机的号数，则其全体基本事件为：

$$A_k=\{k\}\quad (k=1,2,\cdots,n),$$

样本空间为

$$\Omega=\{1,2,\cdots,n\}.$$

除了上述基本事件外，还有如下事件：

$$A_{k,s}=A_k\cup A_s\ (1\leqslant k<s\leqslant n),$$

$$A_{i,k,s}=A_i\cup A_k\cup A_s\ (1\leqslant i<k<s\leqslant n),$$

$$\cdots\cdots$$

$$A_{i_1,i_2,\cdots,i_{n-1}}=A_{i_1}\cup A_{i_2}\cup\cdots\cup A_{i_{n-1}}$$
$$(1\leqslant i_1<i_2<\cdots<i_{n-1}\leqslant n).$$

可以验证，集合族

$$\{\varnothing,\Omega,A_k,A_{k,s},\cdots,A_{i_1,i_2,\cdots,i_{n-1}}\}$$

组成一个 σ-代数．

（2）若只考虑电视机是好品还是次品，则此时基本事件可为 $A_1=$\{好品\}，$A_2=$\{次品\}．可以验证 $\mathscr{F}=\{\varnothing,A_1,A_2,\Omega\}$ 为一个 σ-代数．

例 1.3.2

（1）如例 1.1.5，由一维实数空间的所有子集，包括 $\varnothing,\Omega$，单点集$\{x\}(x\in R_1)$，一切开的、闭的、半开闭区间等组成的集合族$\mathscr{F}$是一 σ-代数．

（2）如例 1.1.5，令

$$A_1=\{x:x\geqslant 0\},$$

$$A_2=\{x: x<0\}.$$

即 A_1 表示"出现非负误差"，A_2 表示"出现负误差"．则可验证

$$\mathscr{F}=\{\varnothing, A_1, A_2, \Omega\}$$

为一代数．

(3) 如例1.1.6，样本空间为

$$\Omega=\{(x,y): 0\leqslant x<\infty, 0\leqslant y<\infty\}.$$

根据定义 1.3.2 后的说明，由坐标平面第一象限内的全体子集组成的集合族是 σ-代数．

从上述两个例子看到，有时虽然研究的具体对象是相同的，但由于所考虑的目的不同，其样本空间和代数或 σ-代数的结构也就不同．特别要指出的是，在这一节以前所说的基本事件也可能不一定是事件，如例 1.3.2 之(2)，基本事件{1}就不属于$\mathscr{F}$．因为$\mathscr{F}$中的事件只有 $\varnothing, A_1, A_2, \Omega$．

由于对于有限样本空间Ω而言，若$\mathscr{F}$是Ω的子集组成的代数，则必为 σ-代数．因此今后我们只讨论 σ-代数，而不单独谈代数．抽去具体对象，我们把任一样本空间 Ω，以及由Ω的子集所组成的一个 σ-代数$\mathscr{F}$写在一起，记为$(\Omega, \mathscr{F})$，称为具有 σ-代数结构的样本空间，或简称为可测空间．特别对有限样本空间，则称为有限可测空间．

可测空间$(\Omega, \mathscr{F})$有如下性质：

性质 1.3.1

1) $\varnothing\in\mathscr{F}$;

2) 若 $A_i\in\mathscr{F}\ (i=1,2,\cdots)$，则

$$\bigcap_{i=1}^{\infty} A_i\in\mathscr{F}.$$

3) 若 $A_i\in\mathscr{F}\ (i=1,2,\cdots,n)$，则

$$\bigcup_{i=1}^{n} A_i \in \mathscr{F},$$

$$\bigcap_{i=1}^{n} A_i \in \mathscr{F}.$$

4）若 $A\in\mathscr{F}$，$B\in\mathscr{F}$，则

$$A-B\in\mathscr{F}.$$

证 1）由定义 1.3.1 之 i）及 ii）得

$$\varnothing=\overline{\Omega}\in\mathscr{F}.$$

2）由 §1.1 可知

$$\bigcap_{i=1}^{\infty} A_i=\overline{\bigcup_{i=1}^{\infty} \overline{A_i}},$$

又由定义 1.3.1 之 ii）及定义 1.3.2 之 iii）′ 得

$$\bigcap_{i=1}^{\infty} A_i \in \mathscr{F}.$$

3）在定义 1.3.2 之 iii）′ 中取

$A_{n+1}=A_{n+2}=\cdots=\varnothing$，则得

$$\bigcup_{i=1}^{n} A_i \in \mathscr{F}.$$

又在2）中取 $A_{n+1}=A_{n+2}=\cdots=\Omega$，则得

$$\bigcap_{i=1}^{n} A_i \in \mathscr{F}.$$

4）因 $A-B=A\cap\overline{B}$，根据定义 1.3.1 之 ii）及上面的 3）得

$$A-B\in\mathscr{F}. \blacksquare$$

在 §1.2 中所述的古典概率及统计概率，所处理的都是有穷样本空间，由古典概率和统计概率的共同性质的揭示，我们可以对有限可测空间定义概率如下：

定义 1.3.3 设 $(\Omega,\mathscr{F})$ 是一个有穷可测空间，对每一集

$A\in\mathscr{F}$，定义实值集函数 $P(A)$，它满足如下三个条件：

i) 对每一 $A\in\mathscr{F}$，有 $0\leqslant P(A)\leqslant 1$；

ii) 对必然事件 Ω，有 $P(\Omega)=1$；

iii) （有穷可加性）设 $A_i\in\mathscr{F}$ $(i=1, 2, \cdots, n)$，$A_i\cap A_j=\varnothing$ $(i\neq j)$，有

$$P\left(\bigcup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i),$$

称 P 为 $(\Omega,\mathscr{F})$ 上的概率，$P(A)$ 称为事件 A 的概率.

而对于像几何概率所处理的非有穷可测空间，还必须满足所谓完全可加性. 对一般可测空间，概率定义如下：

定义1.3.4 设 $(\Omega,\mathscr{F})$ 是可测空间，对每一集 $A\in\mathscr{F}$，有一实数与之对应，记为 $P(A)$（因此在 $\mathscr{F}$ 上定义了一个集函数 P），如它满足下面三个条件

i) 对每一 $A\in\mathscr{F}$，有 $0\leqslant P(A)\leqslant 1$；

ii) 对必然事件 Ω，有 $P(\Omega)=1$；

iii) （完全可加性）对任意 $A_i\in\mathscr{F}(i=1,2, \cdots)$，$A_i\cap A_j=\varnothing$ $(i\neq j)$，恒有

$$P\left(\bigcup_{i=1}^{\infty} A_i\right)=\sum_{i=1}^{\infty} P(A_i),$$

则称实值集函数 P 为 $(\Omega, \mathscr{F})$ 上的概率，$P(A)$ 就称为事件 A 的概率.

例1.3.3 任意抛掷一个质地均匀的硬币，联系这一试验的样本空间为 $\Omega=\{\omega_1,\omega_2\}$，其中 $\omega_1=$（正面），$\omega_2=$（背面），其基本事件为 $A_1=\{\omega_1\}$，$A_2=\{\omega_2\}$. 已知 $\mathscr{F}=\{\varnothing, A_1, A_2, \Omega\}$ 是一个代数，这里 $\varnothing$ 表示"既不出现正面，也不出现背面"这一事件，Ω 表示"出现正面或出现背面"这一事件.

对于 $\mathscr{F}$ 中的集合，对应一实数 P，其定义如下：

$$P(\varnothing)=0;\ P(A_1)=P(A_2)=\frac{1}{2};P(\Omega)=1.$$

显然，P 满足定义 1.3.3，因此 P 是上述有限可测空间$(\Omega,\ \mathscr{F})$上的概率.

例1.3.4 设 $\Omega=\{0,1,2,\cdots\}$，$\mathscr{F}$ 是一切可能的子集所成的集合族，可知$\mathscr{F}$是Ω中的一个 σ-代数. 对任意 $A\in\mathscr{F}$，定义

$$P(A)=\sum_{k\in A}e^{-\lambda}\frac{\lambda^k}{k!}\qquad (\lambda\text{ 为正的常数});$$

$$P(\varnothing)=0. \tag{1.3.1}$$

可以证明它是可测空间$(\Omega,\ \mathscr{F})$上的概率. 我们验证它满足定义 1.3.4.

由于(1.3.1)式，有

$$P(\Omega)=\sum_{k\in\Omega}e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}$$

$$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1.$$

又因为对每一 k

$$e^{-\lambda}\frac{\lambda^k}{k!}\geqslant 0,$$

故

$$0\leqslant P(A)=\sum_{k\in A}e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\leqslant\sum_{k\in\Omega}e^{-\lambda}\frac{\lambda^k}{k!}=1.$$

下面证它满足定义 1.3.4 的 iii).

令 $A_n\in\mathscr{F}(n=1,2,\ \cdots)$，$A_i\cap A_k=\varnothing(i\neq k)$，依照 (1.3.1) 式有

$$P\left(\bigcup_{n=1}^{\infty} A_n\right)=\sum_{k\in\bigcup_{n=1}^{\infty} A_n} e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{n=1}^{\infty}\sum_{k\in A_n}e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{n=1}^{\infty}P(A_n).$$

由下面要讲的性质 1.3.3 可知由概率的完全可加性可推出概率的有穷可加性，因此，定义 1.3.4 确实包括了定义 1.3.3.

这样，无论对古典概率、统计概率或几何概率，它们都满足定义 1.3.4. 因此，根据这一公理化结构，所推出的任何规律，对它们都是适用的.

设 Ω 是一样本空间，$\mathscr{F}$ 为 Ω 中的 σ-代数，P 为 $\mathscr{F}$ 上的概率. 我们称具有上述结构的样本空间为概率空间，记为 $(\Omega, \mathscr{F}, P)$. 例 1.3.3 及例 1.3.4 就是两个概率空间的例子.

设 $(\Omega, \mathscr{F}, P)$ 为概率空间，则概率 P 有如下性质.

性质 1.3.2

$$P\{\varnothing\}=0.$$

证 因为 $\varnothing=\varnothing\cup\varnothing\cup\cdots$，由概率的完全可加性得

$$P\{\varnothing\}=P\{\varnothing\}+P\{\varnothing\}+\cdots$$

因而得

$$P\{\varnothing\}=0. \blacksquare$$

这性质说明：不可能事件的概率为 0. 但逆命题不一定成立，我们将在第二章加以说明.

性质 1.3.3 若 $A_i\in\mathscr{F}$, $i=1,2,\cdots,n$, 且 $A_i\cap A_j=\varnothing\ (i\neq j)$, 则

$$P\left\{\bigcup_{i=1}^{n} A_i\right\}=\sum_{i=1}^{n}P\{A_i\}. \tag{1.3.2}$$

证 由定义 1.3.4 之 iii)，令

$$A_{n+1}=A_{n+2}=\cdots=\varnothing,$$

应用 $P\{\varnothing\}=0$ 这一结果，得

$$P\left\{\bigcup_{i=1}^{n} A_i\right\}=P\left\{\bigcup_{i=1}^{\infty} A_i\right\}=\sum_{i=1}^{\infty}P\{A_i\}=\sum_{i=1}^{n}P\{A_i\}. \blacksquare$$

这一性质说明，由概率的完全可加性可以推得概率的有穷可加性，但逆命题不一定成立.

推论 1) 对任一事件 A，有

$$P\{\bar{A}\}=1-P\{A\}.$$

2) 若 $A\supset B$，则

$$P\{A-B\}=P\{A\}-P\{B\},\text{且}$$

$$P\{A\}\geqslant P\{B\}.$$

证 1) 因为 $A\cup\bar{A}=\Omega$ 且 $A\cap\bar{A}=\varnothing$，故由性质 1.3.3 得

$$P\{A\}+P\{\bar{A}\}=1,$$

移项即得

$$P\{\bar{A}\}=1-P\{A\}.$$

2) 因为 $A=(A-B)\cup B$ 且 $B\cap(A-B)=\varnothing$，故由性质 1.3.3 得

$$P\{A\}=P\{A-B\}+P\{B\},$$

移项得

$$P\{A-B\}=P\{A\}-P\{B\}.$$

又由于 $P\{A-B\}\geqslant 0$，因而得

$$P\{A\}\geqslant P\{B\}. \blacksquare$$

2) 的证明方法具有典型意义，我们可用图示法来说明其证明思想. 如图1.3.1，从图中看到，事件 A 分成两个不相交的部分，一部分为 B，另一部分为图示阴影部分，表示 $A-B$.

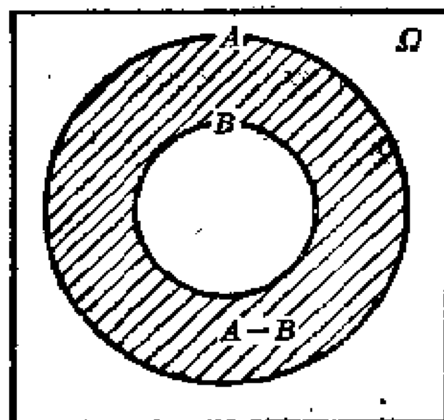


图 1.3.1

推论 1)不仅在计算事件的概率时非常有用，而且在今后一些

定理的证明或公式的推导过程也用到这一性质.

例如在例1.2.5中计算事件"两人有共同生日"的概率 q，可通过事件"两人不同生日"的概率 P 而由推论1)求得：

$$q=1-P\approx1-\left(1-\frac{2}{730}\right)=\frac{2}{730}\approx0.003.$$

性质 1.3.4 （概率连续性）

设 $A_1\supset A_2\supset\cdots$，且 $\bigcap_{n=1}^{\infty}A_n=\varnothing$，

则

$$P(A_n)\to0\quad(n\to\infty).$$

证 由假设，可得

$$A_n=\bigcup_{k=n}^{\infty}(A_k-A_{k+1})\quad(n=1,2,\cdots),$$

而 $A_k-A_{k+1}(k=1,2,\cdots)$ 互不相容，如图1.3.2. 由完全可加性得

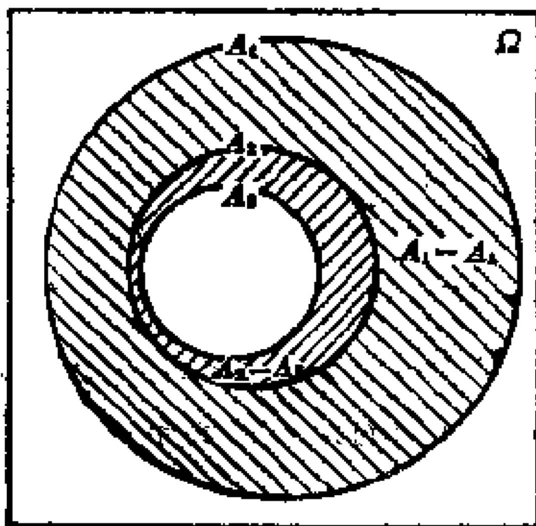


图 1.3.2

$$1\geqslant P(A_1)=P\left\{\bigcup_{k=1}^{\infty}(A_k-A_{k+1})\right\}=\sum_{k=1}^{\infty}P(A_k-A_{k+1}).$$

由于 $P(A_n)=\sum_{k=n}^{\infty}P(A_k-A_{k+1})$ 是上面收敛级数的尾项，因而

$$P(A_n) \to 0 \quad (n \to \infty). \quad \blacksquare$$

推论 若 $A_1 \supset A_2 \supset \cdots$，且 $\bigcap_{n=1}^{\infty} A_n = A$，则

$$P(A_n) \to P(A) \quad (n \to \infty).$$

事实上，令 $B_n = A_n - A$，则 $B_1 \supset B_2 \supset \cdots$，且

$$\bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1}^{\infty} (A_n \cap \bar{A}) = \left(\bigcap_{n=1}^{\infty} A_n\right) \cap \bar{A} = A \cap \bar{A} = \varnothing.$$

由性质 1.3.4 得 $P(B_n) \to 0$，即

$$P(A_n) - P(A) = P(A_n - A) = P(B_n) \to 0.$$

这就得到了

$$P(A_n) \to P(A) \quad (n \to \infty). \quad \blacksquare$$

一般称具有推论所述性质的概率 P 为连续的。读者可自行证明，若 $A_1 \subset A_2 \subset \cdots$，且

$$\bigcup_{n=1}^{\infty} A_n = A,$$

则

$$P(A_n) \to P(A) \quad (n \to \infty).$$

从性质 1.3.3 及性质 1.3.4 我们得到：从概率的完全可加性可以推出概率的有限可加性和概率的连续性. 可以证明，其逆命题也是成立的，我们有如下定理.

定理 1.3.1 设 P 为可测空间 $(\Omega, \mathscr{F})$ 上的非负实值集函数，$P(\Omega) = 1$，则 P 为完全可加的充分必要条件为

1) P 是有穷可加的；

2) P 是连续的.

证 必要性由性质 1.3.3 及性质 1.3.4 已证. 下面证充分性. 设 $A_1, A_2, \cdots$ 均属于 $\mathscr{F}$，$A_i \cap A_k = \varnothing$ $(i \neq k)$. 令

$$A = \bigcup_{i=1}^{\infty} A_i, \quad B_n = \bigcup_{i=1}^{n} A_i, \quad C_n = A - B_n,$$

则 A, B_n, C_n(对一切 n)均属于$\mathscr{F}$，且 $A\supset B_n$，同时，

$$C_1\supset C_2\supset\cdots,\quad \bigcap_{i=1}^{\infty} C_n=\varnothing.$$

由概率连续性假设，有 $P(C_n)\to 0\ (n\to\infty)$，即

$$P(A)-P(B_n)=P(C_n)\to 0\ (n\to\infty).$$

另一方面，由有穷可加性得

$$P(B_n)=\sum_{i=1}^{n}P(A_i).$$

即得

$$P(A)=\sum_{i=1}^{\infty}P(A_i).\ \blacksquare$$

性质 1.3.5 (多除少补原理). 设 $A_i\in\mathscr{F}\ (i=1,2,\cdots)$. 则

$$P\left(\bigcup_{i=1}^{n}A_i\right)=S_1-S_2+S_3-S_4+\cdots+(-1)^{n+1}S_n, \qquad (1.3.3)$$

其中

$$S_1=\sum_{i=1}^{n}P(A_i);$$

$$S_2=\sum_{1\leqslant i<k\leqslant n}P(A_i\cap A_k);$$

$$S_3=\sum_{1\leqslant i<k<s\leqslant n}P(A_i\cap A_k\cap A_s);$$

$$\cdots\cdots\cdots\cdots$$

$$S_n=P(A_1\cap A_2\cap\cdots\cap A_n).$$

证 用归纳法. 设 $A_1\in\mathscr{F}$, $A_2\in\mathscr{F}$. 由于 $A_1\cup A_2=A_1\cup[A_2-(A_1\cap A_2)]$(如图 1.3.3)，且 A_1 与$[A_2-(A_1\cap A_2)]$互不相容，故得

$$P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1\cap A_2). \qquad (1.3.4)$$

设 $n=k$ 时(1.3.3)式成立，因为

$$P\left(\bigcup_{i=1}^{k+1} A_i\right)=P\left[\left(\bigcup_{i=1}^{k} A_i\right)\cup A_{k+1}\right],$$

由(1.3.4)式得

$$P\left(\bigcup_{i=1}^{k+1} A_i\right)=P\left(\bigcup_{i=1}^{k} A_i\right)+P(A_{k+1})-P\left\{\left(\bigcup_{i=1}^{k} A_i\right)\cap A_{k+1}\right\}. \tag{1.3.5}$$

因设 $n=k$ 时(1.3.3)式成立，即

$$P\left(\bigcup_{i=1}^{k} A_i\right)=S_1-S_2+S_3-S_4+\cdots+(-1)^{k+1}S_k,$$

其中

$$S_1=\sum_{i=1}^{k}P(A_i);$$

$$S_2=\sum_{1\leqslant i<r\leqslant k}P(A_i\cap A_r);$$

$$S_3=\sum_{1\leqslant i<r<s\leqslant k}P(A_i\cap A_r\cap A_s);$$

$$\cdots\cdots\cdots\cdots$$

$$S_k=P(A_1\cap A_2\cap\cdots\cap A_k).$$

同时，

$$P\left[\left(\bigcup_{i=1}^{k} A_i\right)\cap A_{k+1}\right]=P\left[\bigcup_{i=1}^{k}(A_i\cap A_{k+1})\right]$$

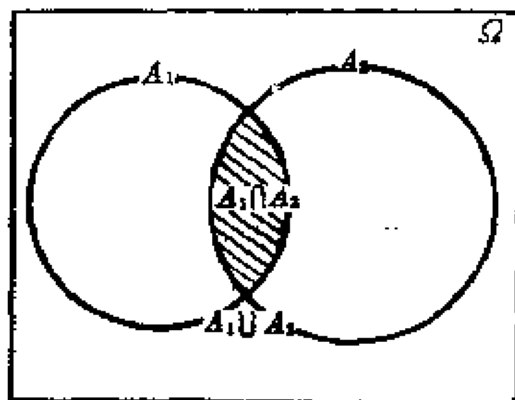


图 1.3.3

$$=S_1'-S_2'+S_3'-S_4'+\cdots+(-1)^{k+1}S_k'$$

其中

$$S_1'=\sum_{i=1}^{k}P(A_i\cap A_{k+1});$$

$$S_2'=\sum_{1\leqslant i<r\leqslant k}P(A_i\cap A_r\cap A_{k+1});$$

$$S_3'=\sum_{1\leqslant i<r<s\leqslant k}P(A_i\cap A_r\cap A_s\cap A_{k+1});$$

$$\cdots\cdots\cdots\cdots$$

$$S_k'=P(A_1\cap A_2\cap\cdots\cap A_k\cap A_{k+1}),$$

而

$$\left.\begin{aligned}
&S_1+P(A_{k+1})=\sum_{i=1}^{k+1}P(A_i);\\
&S_2+S_1'=\sum_{1\leqslant i<r\leqslant k+1}P(A_i\cap A_r);\\
&S_3+S_2'=\sum_{1\leqslant i<r<s\leqslant k+1}P(A_i\cap A_r\cap A_s);\\
&\cdots\cdots\cdots\cdots\\
&S_k'=P(A_1\cap A_2\cap\cdots\cap A_{k+1}).
\end{aligned}\right\}\quad(1.3.6)$$

根据（1.3.5）式知（1.3.3）式对 $n=k+1$ 也成立，因此得所证结果. ∎

推论（次可加性）

设 $A_i\in\mathscr{F}$ $(i=1,2,\cdots,n)$，则有

$$P\left(\bigcup_{i=1}^{n}A_i\right)\leqslant\sum_{i=1}^{n}P(A_i).\qquad(1.3.7)$$

证 用归纳法. 当 $n=2$ 时，由(1.3.4)式立即可得

$$P(A_1\cup A_2)\leqslant P(A_1)+P(A_2).$$

若 $n=k$ 时(1.3.7)式成立，则由(1.3.5)式可得，当 $n=k+1$ 时有

$$P\left(\bigcup_{i=1}^{k+1} A_i\right) \leqslant \sum_{i=1}^{k} P(A_i) + P(A_{k+1}) = \sum_{i=1}^{k+1} P(A_i). \blacksquare$$

最后，我们指出，在定义 1.3.3 或定义 1.3.4 中，概率 P 要求对一切 $A \in \mathscr{F}$ 都有定义. 有时候，只须对 $\mathscr{F}$ 中的某一部分有定义即可，其余利用概率 P 的性质随之也可以确定. 例如，对古典概率的概率空间只须给定其基本事件的概率为 $\frac{1}{n}$ 即 $P\{\omega_i\} = \frac{1}{n}$ $(i=1,2,\cdots,n)$，其中 $\{\omega_i\}$ $(i=1,2,\cdots,n)$ 为全体基本事件. 对于其它事件 $A \in \mathscr{F}$ 的概率，则可通过概率的可加性得到. 如事件 A 是由 m 个基本事件 $A_{i_k} = \{\omega_{i_k}\}$ $(k=1,2,\cdots,m, m \leqslant n)$ 组成，按概率可加性

$$P(A) = P\left(\bigcup_{k=1}^{m} A_{i_k}\right) = \sum_{k=1}^{m} P(A_{i_k}) = m \cdot \frac{1}{n} = \frac{m}{n}.$$

这正是我们在 § 1.2 中计算古典概率的公式.

§ 1.4 条件概率

一、条件概率的定义、例及性质

在实际问题中，一般除了要考虑事件 A 的概率 $P(A)$，还须考虑在"已知事件 B 已发生"这一条件下，事件 A 发生的概率. 一般地说，后者的概率与前者的概率未必相同. 为了区别起见，我们把后者叫作条件概率，记为 $P(A|B)$ 或 $P_B(A)$，读作在条件 B 下，事件 A 的条件概率.

例 1.4.1 设 100件某一产品中有 5 件不合格品，而 5 件不合格品中又有 3 件是次品，2件是废品. 现任意在100件产品中抽取一件(设 100 件产品都以同等可能被抽到)，求：

1) 抽得的是废品的概率；

2) 已知抽得的是不合格品，它是废品的概率.

解 令A表示"抽得的是废品"这一事件，B表示"抽得的是不合格品"这一事件.

1) 由于100件产品中有2件是废品，按古典概率计算得

$$P(A)=\frac{2}{100}=\frac{1}{50}.$$

2) 由于5件不合格品中，废品只有2件，故

$$P(A|B)=\frac{2}{5}.$$

由此看到

$$P(A)\neq P(A|B).$$

一般地，条件概率如何定义呢？我们先从分析例1.4.1着手：计算$P(B)$及$P(A\cap B)$.

由于100件产品中有5件不合格品，故$P(B)=\frac{5}{100}$. $P(A\cap B)$表示"抽得的是不合格品且是废品"这一事件的概率，而由于不合格品且是废品的个数在100个中有2个，故$P(A\cap B)=\frac{2}{100}$，通过简单的运算得

$$P(A|B)=\frac{2}{5}=\frac{\frac{2}{100}}{\frac{5}{100}}=\frac{P(A\cap B)}{P(B)}. \qquad (1.4.1)$$

由(1.4.1)式的启发，我们对条件概率$P(A|B)$定义如下：

定义1.4.1 设$(\Omega,\mathscr{F},P)$为一概率空间，$A\in\mathscr{F}$，$B\in\mathscr{F}$，且$P(B)>0$，在"已知事件B已经发生"条件下，"事件A发生"的条件概率$P(A|B)$定义为

$$P(A|B)=\frac{P(A\cap B)}{P(B)}. \qquad (1.4.2)$$

例1.4.2 任意抛掷两次质地均匀的硬币，令B表示"第一

次出现正面"，A 表示"第二次出现正面"这两个事件，求 $P(A|B)$.

解 抛掷两次硬币的全体基本事件为：{(正面，正面)}，{(正面，背面)}，{(背面，正面)}，{(背面，背面)}. 事件 $A\cap B$ 为上面四个基本事件之第一个，故

$$P(A\cap B)=\frac{1}{4}.$$

同理可证 $P(B)=\frac{2}{4}=\frac{1}{2}$，由(1. 4.2)式得

$$P(A|B)=\frac{1}{4}\div\frac{2}{4}=\frac{1}{2}.$$

这一结果，可由直接分析全体基本事件，按古典概率直接计算加以验证：因为在四个基本事件中，"第一次出现正面"的有两个，而在这两个中"第二次出现正面"的有一个，故 $P(A|B)=\frac{1}{2}$.

从条件概率的定义并不能得出事件 A 的概率 $P(A)$ 与条件概率 $P(A|B)$ 有什么必然关系，如不能说 $P(A)\leqslant P(A|B)$ 或 $P(A|B)\leqslant P(A)$. 例如从 0, 1, …, 9 十个数字中，任意抽取一数，A 表示"抽得一数为 3 的倍数"，B 表示"抽得一数为偶数"，B_1 表示"抽得一数大于 8"；则 $P(A)=\frac{3}{10}$，$P(A|B)=\frac{1}{5}$，$P(A|B_1)=1$. 因此 $P(A)>P(A|B)$，$P(A)<P(A|B_1)$.

当 $B=\Omega$ 时，由于 $A\cap\Omega=A$，因此

$$P(A|\Omega)=\frac{P(A\cap\Omega)}{P(\Omega)}=\frac{P(A)}{1}=P(A).$$

从直观意义看，上式的结果是明显的. 因为 Ω 是必然事件，因此，在必然事件发生这个条件下考虑事件 A 的概率与事件 A 的无条件概率应是一样的.

对任意事件 A 与事件 B 在下述三种特殊情况下，条件概率 $P(A|B)$ 与 $P(A)$ 有如下关系：

1) 若 $A\subset B$，则 $P(A|B)\geqslant P(A)$；

事实上，由于 $A\cap B=A$，故

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)}{P(B)}\geqslant P(A).$$

2）若 $B\subset A$，则 $P(A|B)\geqslant P(A)$；

事实上，由 $B\subset A$ 知，当 B 发生时，A 必然发生，故 $P(A|B)=1$，从而 $P(A)\leqslant 1=P(A|B)$.

3）若 $A\cap B=\varnothing$，即 A 与 B 互不相容，则

$$P(A|B)\leqslant P(A).$$

事实上，因 A 与 B 互不相容，故 B 发生则 A 必然不发生，从而 $P(A|B)=0$，由此得 $P(A)\geqslant 0=P(A|B)$.

条件概率有如下性质：

性质 1.4.1 设 $(\Omega,\mathscr{F},P)$ 为一概率空间，$B\in\mathscr{F}$ 且 $P(B)>0$；则对任意 $A\in\mathscr{F}$，有 $P(A|B)$ 对应，且 $P(A|B)$ 是 $(\Omega,\mathscr{F})$ 上的概率，即 $P(A|B)$ 满足

1）$0\leqslant P(A|B)\leqslant 1$；

2）$P(\Omega|B)=1$；

3）若 $A_i\in\mathscr{F}(i=1,2,\cdots)$ 且 $A_i\cap A_k=\varnothing(i\neq k)$，则

$$P\left(\bigcup_{i=1}^{\infty}A_i\,\middle|\,B\right)=\sum_{i=1}^{\infty}P(A_i|B).$$

证 1）$\because\ A\cap B\subset B$，

$$\therefore\quad P(A\cap B)\leqslant P(B).$$

因此
$$P(A|B)=\frac{P(A\cap B)}{P(B)}\leqslant 1.$$

$P(A|B)\geqslant 0$ 是显然的.

2）$\because\quad \Omega\cap B=B$，

$$\therefore\quad P(\Omega|B)=\frac{P(\Omega\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1.$$

3）$\because\ \left(\bigcup_{i=1}^{\infty}A_i\right)\cap B=\bigcup_{i=1}^{\infty}(A_i\cap B)$，

并由 $i\neq k$ 时 $A_i\cap A_k=\varnothing$，得

$$(A_i\cap B)\cap(A_k\cap B)\subset A_i\cap A_k=\varnothing,$$

即

$$(A_i\cap B)\cap(A_k\cap B)=\varnothing \qquad (i\neq k).$$

由此得到

$$P\left(\bigcup_{i=1}^{\infty}A_i\,\middle|\,B\right)=\frac{P\left[\left(\bigcup_{i=1}^{\infty}A_i\right)\cap B\right]}{P(B)}=\frac{P\left[\bigcup_{i=1}^{\infty}(A_i\cap B)\right]}{P(B)}$$

$$=\sum_{i=1}^{\infty}\frac{P(A_i\cap B)}{P(B)}=\sum_{i=1}^{\infty}P(A_i|B). \quad \blacksquare$$

如 $P(B)>0$，且用 P_B 表示在"事件 B 发生"条件下的条件概率，这样，性质 1.4.1 说明 $(\Omega,\mathscr{F},P_B)$ 也是一个概率空间，称它为条件概率空间. 因此 §1.3 中对概率所证明的性质，对于条件概率而言，同样具有那些性质，这里不再重述.

由上面所述，我们可以从两种不同的观点来考虑条件概率：

一种观点是可测空间 $(\Omega,\mathscr{F})$ 没有变，其概率变了，即 P 换为 P_B，它对任意 $A\in\mathscr{F}$ 都有定义，并且

$$P_B(A)=\frac{P(A\cap B)}{P(B)}. \tag{1.4.3}$$

概率空间由 $(\Omega,\mathscr{F},P)$ 变为 $(\Omega,\mathscr{F},P_B)$.

另一种观点是可测空间变了，Ω 变为 $\Omega_1=\Omega\cap B$，而 $\mathscr{F}$ 变为 $\mathscr{F}_1=\{C\cap B: C\in\mathscr{F}\}$，这样，$\Omega_1$ 中的样本点缩小了，它是由 Ω 中除去不在 B 中的样本点组成的. 于是我们得到了另一个概率空间 $(\Omega_1,\mathscr{F}_1,P'_B)$，其中 P'_B 的定义为：对任意 $A\in\mathscr{F}_1$

$$P'_B(A)=\frac{P(A)}{P(B)}. \tag{1.4.3$'$}$$

这两种观点其实没有什么不同，前一观点较简单，它把样本空间缩小了这一事实掩盖了，把应当除去的样本点保留了下来. 然

而，对Ω中的任何A，Ω中的$A\cap B$就是Ω_1中的A，所以(1.4.3)和(1.4.3)′的右端并无不同.

性质 1.4.2 设A为概率空间$(\Omega,\mathscr{F},P)$上的正概率事件，又设$B\in\mathscr{F}$，且$P_A(B)>0$，则对条件概率空间$(\Omega,\mathscr{F},P_A)$而言，对任意$C\in\mathscr{F}$有

$$P_A(C|B)=P(C|A\cap B).$$

证

$$P_A(C|B)=\frac{P_A(B\cap C)}{P_A(B)}$$

$$=\frac{P(A\cap B\cap C)}{P(A)}\div\frac{P(A\cap B)}{P(A)}$$

$$=\frac{P(A\cap B\cap C)}{P(A\cap B)}=P(C|A\cap B).\ \blacksquare$$

性质1.4.2 说明计算条件概率空间中任一事件的条件概率，可化为在原来的概率空间中求该事件的条件概率.

二、乘法公式

由条件概率定义，可得，

$$P(A\cap B)=P(A|B)\cdot P(B);\tag{1.4.4}$$

$$P(A\cap B)=P(B|A)\cdot P(A).\tag{1.4.5}$$

上面(1.4.4)，(1.4.5)式均称为事件概率的乘法公式.

这里我们要注意，式(1.4.4)，(1.4.5)同时成立，是在条件$P(A)>0$，且$P(B)>0$之下而言的．若$P(A)$或$P(B)$之中有一个等于0，则不能同时成立．例如$P(A)=0$，则只有乘法公式(1.4.4)成立，即

$$P(A\cap B)=P(A|B)P(B),$$

而另一个乘法公式

$$P(A\cap B)=P(B|A)P(A)$$

不成立，这是因为$P(B|A)$无定义.

上述乘法公式可推广到任意有穷多个事件时的情形：

性质1.4.3 设 $A_i \in \mathscr{F}(i=1,2,\cdots,n,\ n\geqslant 2)$ 满足 $P(A_1\cap A_2\cap\cdots\cap A_{n-1})>0$，则

$$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1\cap A_2)\cdots\cdot P(A_n|A_1\cap A_2\cap\cdots\cap A_{n-1}) \tag{1.4.6}$$

证 $\because P(A_1)\geqslant P(A_1\cap A_2)\geqslant\cdots\geqslant P(A_1\cap A_2\cdots\cap A_{n-1})>0$

故(1.4.6)右端有定义，且为

$$P(A_1)\cdot\frac{P(A_1\cap A_2)}{P(A_1)}\cdot\frac{P(A_1\cap A_2\cap A_3)}{P(A_1\cap A_2)}\cdots\cdots\frac{P(A_1\cap A_2\cap\cdots\cap A_n)}{P(A_1\cap A_2\cap\cdots\cap A_{n-1})}$$

$$=P(A_1\cap A_2\cap\cdots\cap A_n).\ ▮$$

例1.4.3 设 100 件产品中有 5 件是不合格品，用下列两种方法抽取 2 件，求 2 件都是合格品的概率：

1) 不放回顺序抽取；

2) 放回顺序抽取；

解 令 $A=${第一次抽得是合格品}，

$B=${第二次抽得是合格品}。

我们的问题是求 $P(A\cap B)$。

1) 由题设，不放回抽取时，

$$P(A)=\frac{95}{100};\quad P(B|A)=\frac{94}{99}.$$

由式(1.4.5)算得

$$P(A\cap B)=\frac{95}{100}\cdot\frac{94}{99}\approx 0.902$$

2) 由题设，用放回抽取时，

$$P(A)=\frac{95}{100};\quad 而\ P(B|A)=\frac{95}{100}.$$

由(1.4.5)式算得

$$P(A\cap B)=\frac{95}{100}\cdot\frac{95}{100}=0.9025.$$

在 2) 的假设下，我们可以求得 $P(B)=\frac{95}{100}$，它正好等于 $P(B|A)$，即 $P(B)=P(B|A)$. 它说明事件 A 发生与否不影响事件 B 发生的概率. 这结论从 2) 的假设可以直接看到，因为此时第二次抽取时的条件与第一次抽取时完全相同，即第一次抽取的结果，完全不影响第二次抽取.

一个事件的发生与否，不影响另一事件发生可能性的大小(即两个事件之间有某种"独立性")这一性质，在概率论里是需要进一步研究的，这一点我们将在§1.5讲述.

例1.4.4 (配对问题)：某人写了 n 封信，将其放入信封中，并在其中每一个信封上分别任意地写上 n 个收信人中的一个地址(不重复). 求：1)没有一个信封上所写的地址正确[①]的概率 q_0；2) 恰有 r 个信封上所写的地址正确的概率 $p_r(r\leqslant n)$.

解 设 A_i 表示"在第 i 个信封上所写的地址正确"这一事件 $(i=1,2,\cdots,n)$.

1) 显然有

$$q_0=1-P\left(\bigcup_{i=1}^{n}A_i\right).$$

由于事件 A_i 是相容的，故需用性质1.3.5(多除少补原理)来计算，并且应用性质1.4.3(乘法公式).

依题意对任意 $i<j<k<\cdots$，有

$$P(A_i)=\frac{1}{n};$$

$$P(A_i\cap A_j)=P(A_i)P(A_j|A_i)=\frac{1}{n}\cdot\frac{1}{n-1}=\frac{(n-2)!}{n!};$$

---

① 所谓"正确"是指信封上所写地址，恰好是该信封内装进的信的收信人的地址.

$$P(A_i \cap A_j \cap A_k) = \frac{(n-3)!}{n!};$$

…………

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \frac{1}{n!}.$$

由此有

$$S_1 = \sum_{i=1}^{n} P(A_i) = 1;$$

$$S_2 = \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) = C_n^2 \frac{(n-2)!}{n!};$$

$$S_3 = \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k)$$

$$= C_n^3 \cdot \frac{(n-3)!}{n!};$$

…………

$$S_n = \frac{1}{n!}.$$

于是

$$q_0 = 1 - P\left(\sum_{i=1}^{n} A_i\right)$$

$$= 1 - \left[1 - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} + \cdots + (-1)^{n+1} \frac{1}{n!}\right]$$

$$= \sum_{k=0}^{n} \frac{(-1)^k}{k!},$$

注意 q_0 与 n 有关，如记 $q_0 = q_0(n)$，则

$$\lim_{n \to \infty} q_0(n) = e^{-1} \approx 0.37.$$

因此当 n 很大时，$q_0 \approx 0.37$.

2）在指定的"某 r 个信封上所写的地址是正确的"这一事件

的概率应为

$$\frac{(n-r)!}{n!}=\frac{1}{n(n-1)\cdots(n-r+1)},$$

而其余的“$n-r$ 个信封上所写的地址均不正确”这事件的概率 $q_0(n-r)$，读者稍加思考，可求得为

$$q_0(n-r)=\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}.$$

由于 r 个信封共有 C_n^r 种选法，故得所求概率为

$$\begin{aligned} p_r&=C_n^r\frac{1}{n(n-1)\cdots(n-r+1)}\cdot\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}\\ &=\frac{1}{r!}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}. \end{aligned}$$

当 $n\to\infty$ 时，$p_r\to\frac{1}{r!}e^{-1}$.

三、全概率公式

每一随机试验，它的全体基本事件可以用各种不同的方法分成若干类(我们说对样本空间进行剖分)，任一复合事件都可用这几类基本事件的复合而得到. 如从 0, 1, 2, …, 9 这十个数字中任取一个数，其结果可分为两类：奇数，偶数. 这样，“大于 5 的数”就可看成是“偶数而大于 5 的数”与“奇数而大于 5 的数”的和.

我们还可以再看一个例子：设 n 个袋子，各装有白球和黑球，任意选取一袋子，并任取一球，则“取出一球是白球”这一事件可由“从第一袋子抽出一球是白球”，“从第二袋子抽出一球是白球”，…，“从第 n 个袋子抽出一球是白球”等复合而成.

上述问题的直观解释是：对一个试验，某一结果的发生可能有多种原因，每一原因对这结果的发生作出一定的“贡献”，当然，这结果发生的可能性与各种原因的“贡献”大小有关.

对于这一类问题，从概率上表达它们发生可能性之间的关系的一个公式就是所谓全概率公式.

在叙述全概率公式以前，我们先引入一个以后有用的概念.

定义1.4.2 设$(\Omega,\mathscr{F},P)$为概率空间，如果$A_i\subset\Omega$ $(i=1,2,\cdots,n)$，$A_i\cap A_j=\varnothing$ $(i\neq j)$，且$\bigcup\limits_{i=1}^{n}A_i=\Omega$，则称$A_1,A_2,\cdots,A_n$为$\Omega$的一个有穷剖分.

如果$A_i\subset\Omega$ $(i=1,2,\cdots)$，$A_i\cap A_j=\varnothing$，且$\bigcup\limits_{i=1}^{\infty}A_i=\Omega$，则称$A_1,A_2,\cdots,A_n,\cdots$为$\Omega$的一个可列无穷剖分. 如图1.4.1是$n=8$的剖分.

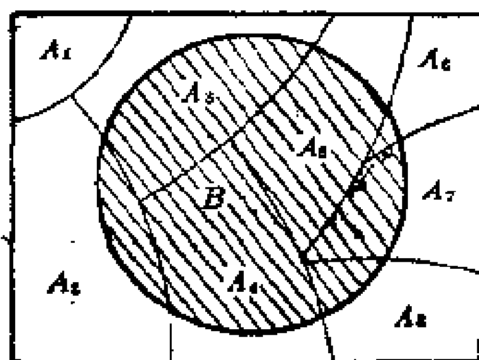


图 1.4.1

定理1.4.1(全概率公式)

设$(\Omega,\mathscr{F},P)$为一概率空间，$A_1,A_2,\cdots,A_n$是Ω的一个有穷剖分，且$P(A_i)>0$ $(i=1,2,\cdots,n)$. 则对任一事件$B\in\mathscr{F}$，有

$$P(B)=\sum_{i=1}^{n}P(B|A_i)P(A_i). \tag{1.4.7}$$

(1.4.7)式称为全概率公式.

证 ∵ $B=B\cap\Omega=B\cap\left(\bigcup\limits_{i=1}^{n}A_i\right)=\bigcup\limits_{i=1}^{n}(B\cap A_i)$.

又由于

$$A_i\cap A_k=\varnothing \quad (i\neq k).$$

故

$$(B\cap A_i)\cap(B\cap A_k)=\varnothing \quad (i\neq k).$$

应用概率的有穷可加性，并注意到(1.4.5)式得

$$P(B)=\sum_{i=1}^{n}P(B\cap A_i)=\sum_{i=1}^{n}P(B|A_i)P(A_i). \blacksquare$$

例1.4.5 设1000件产品中有200件是不合格产品，依次作不放回抽取二件产品，求第二次取到的是不合格品的概率.

解 令 $A=\{$第一次取到的是不合格品$\}$

$B=\{$第二次取到的是不合格品$\}$

记 $A_1=A$, $A_2=\bar{A}$, 用全概率公式:

$$P(B)=P(B|A)P(A)+P(B|\bar{A})P(\bar{A}).$$

由题意

$$P(A)=\frac{200}{1000}=\frac{1}{5}.$$

$$\therefore \quad P(\bar{A})=1-P(A)=\frac{4}{5}.$$

由于是不放回抽取，所以

$$P(B|A)=\frac{199}{999}, \quad P(B|\bar{A})=\frac{200}{999},$$

故有

$$P(B)=\frac{199}{999}\times\frac{1}{5}+\frac{200}{999}\times\frac{4}{5}=\frac{1}{5}.$$

定理1.4.1(全概率公式)对于可列无穷剖分，且 $P(A_i)>0$ 的情形，也是成立的，即对任意 $B\in\mathscr{F}$ 有

$$P(B)=\sum_{i=1}^{\infty}P(B|A_i)P(A_i). \tag{1.4.8}$$

事实上，只要在定理1.4.1的证明过程中将概率有穷可加性换为概率完全可加性即可.

定理1.4.1中的条件可改为:

设 $A_1, A_2, \cdots$ 为有穷或可列无穷个互不相容的事件，$B\in\mathscr{F}$ 且 $\bigcup_i A_i\supset B$，若 $P(A_i)>0(i=1,2,\cdots)$，则有

$$P(B)=\sum_i P(B|A_i)P(A_i). \tag{1.4.9}$$

事实上，由条件知

$$B=\left(\bigcup_i A_i\right)\cap B=\bigcup_i (A_i\cap B),$$

且

$$(A_i\cap B)\cap(A_k\cap B)=\varnothing\quad(i\neq k).$$

再由有穷可加性或完全可加性及(1.4.5)式得

$$P(B)=\sum_i P(A_i\cap B)=\sum_i P(B|A_i)P(A_i).$$

推论 设$(\Omega,\mathscr{F},P)$为一概率空间，$A_i\in\mathscr{F}$是任一有穷或可列无穷剖分，则对固定的事件$D\in\mathscr{F}$，对任一$A_i\in\mathscr{F}$，$P(A_i\cap D)>0$以及任意$B\in\mathscr{F}$有

$$P(B|D)=\sum_i P(A_i|D)P(B|A_i\cap D).\qquad(1.4.10)$$

公式(1.4.10)称为条件全概率公式.

事实上由性质1.4.2及全概率公式得

$$\begin{aligned}P(B|D)&=P_D(B)\\&=\sum_i P_D(A_i)P_D(B|A_i)\\&=\sum_i P(A_i|D)P(B|A_i\cap D).\end{aligned}$$

四、贝叶斯公式

全概率公式给了我们一个实际计算某些事件概率的公式，只要一旦我们知道了在各事件(注意这些事件应满足定义1.4.2关于剖分的条件)发生条件下该事件发生的概率，则该事件的无条件概率可从全概率公式求得. 直观地说，只要知道了各种原因发生条件下该事件发生的概率(姑且叫它"原因"概率)，该事件的无条件概率可通过全概率公式求得.

上述问题的"逆问题"可叙述如下：

若已知各种"原因"概率，设在进行随机试验中该事件已发生，问在这条件下，各原因发生的条件概率是多少？例如前述关于抽

取白球的例子里，若已知抽出的一球是白球，问它从哪一袋取出的可能性最大？又如在诊病问题中，若我们从病理或长期积累的经验中，知道了有多种病因(这里假定这些病因满足定义 1.4.2 剖分要求，即它们之间的不相容性)会产生某症状，并且知道这些"原因"概率. 假若在一次诊病的病例中已经出现该症状，问其最大可能的原因是什么？类似这样的例子，在理论和实际中经常见到，解决这类问题的公式就是下面讲述的贝叶斯公式.

定理1.4.2 设$(\Omega,\mathscr{F},P)$为概率空间，$A_i\in\mathscr{F}(i=1,2,\cdots,n)$为$\Omega$的一个有穷剖分且$P(A_i)>0(i=1,2,\cdots,n)$，则对任意$B\in\mathscr{F}$且$P(B)>0$，有

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)}(i=1,2,\cdots,n). \qquad (1.4.11)$$

(1.4.11)式称为贝叶斯公式.

证 由条件概率及全概率公式得

$$P(A_i|B)=\frac{P(A_i\cap B)}{P(B)}=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)}.$$

定理 1.4.2 中的有穷数 n 可以推广到可列无穷的情形，即贝叶斯　　可写为

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty}P(B|A_j)P(A_j)} \qquad (i=1,2,\cdots). \quad (1.4.12)$$

在定理 1.4.1 下面的小字部分所述的条件下，(1.4.11)及(1.4.12)同样成立.

有时，我们把$P(B|A_i)$称为"原因"概率，而称$P(A_i|B)$为"事后"概率. 贝叶斯公式告诉我们："事后"概率可通过一系列的"原因"概率求得.

例1.4.6 设8支枪中有3支未经试射校正，5支已经试射校正. 一射击手用校正过的枪射击时，中靶概率为0.8；而用未校正过的枪射击时，中靶概率为0.3. 今假定从8支枪中任取一支进行射击，结果中靶，求所用这支枪是已校正过的概率.

解 设 A_1 和 A_2 分别表示"所取的枪是校正过的"和"所取的枪是未校正过的"这两事件，B 表示"射击中靶"这一事件.
由题设

$$P(A_1)=\frac{5}{8};\quad P(A_2)=\frac{3}{8};$$

$$P(B|A_1)=0.8;P(B|A_2)=0.3.$$

于是依贝叶斯公式得所求的概率为

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)}$$

$$=\frac{\frac{5}{8}\times 0.8}{\frac{5}{8}\times 0.8+\frac{3}{8}\times 0.3}=\frac{4}{4.9}\approx 0.82.$$

例1.4.7 设有编号为(I)、(II)、(III)的三个箱子，(I)箱内装有 n_1 个白球和 m_1 个黑球，(II)箱内装有 n_2 个白球和 m_2 个黑球，(III)箱内装有 n_3 个白球和 m_3 个黑球，今任意取出一箱，再自此箱中任取一球(每一箱或每一球均设具有等可能被抽取到)，结果发现为白球. 试求在事件"此球为白球"(记为 B)的条件下，事件"此球属于(I)箱"(记为 A_1)的条件概率 $P(A_1|B)$.

解 令 A_2 = {此球为属于(II)箱}，A_3 = {此球为属于(III)箱}，由题意知

$$P(A_1)=P(A_2)=P(A_3)=\frac{1}{3}>0,$$

A_1, A_2, A_3 互不相容且 $\bigcup_{i=1}^{3} A_i = \Omega$.

由全概率公式得

$$P(B) = \sum_{i=1}^{3} P(B|A_i)P(A_i)$$

$$= \frac{1}{3}\left[\frac{n_1}{n_1+m_1} + \frac{n_2}{n_2+m_2} + \frac{n_3}{n_3+m_3}\right] > 0.$$

再由贝叶斯公式,得

$$P(A_1|B) = \frac{\frac{1}{3}\left(\frac{n_1}{n_1+m_1}\right)}{\frac{1}{3}\left[\frac{n_1}{n_1+m_1} + \frac{n_2}{n_2+m_2} + \frac{n_3}{n_3+m_3}\right]}$$

$$= \frac{1}{1 + \frac{n_2(n_1+m_1)}{n_1(n_2+m_2)} + \frac{n_3(n_1+m_1)}{n_1(n_3+m_3)}}.$$

例 1.4.8 在数字通讯中,信号是由数字 0 和 1 的长序列组成的,由于有随机干扰,发送的信号 0 或 1 各有可能错误接收为 1 或 0.现假定发送信号为 0 和 1 的概率均为 $\frac{1}{2}$,又已知发送 0 时,接收为 0 和 1 的概率分别为 0.8 和 0.2;发送信号为 1 时,接收为 1 和 0 的概率分别为 0.9 和 0.1;求:已知收到信号是 0 时,发出的信号是 0 (即没有错误接收)的概率.

解 令 A_{i+1} = {发出信号是 i} $(i=0,1)$;

B = {收到信号是 0},

由假设知

$$P(A_1) = P(A_2) = \frac{1}{2}; P(B|A_1) = 0.8;$$

$$P(B|A_2) = 0.1.$$

由贝叶斯公式得所求的概率为

$$P(A_1|B)=\frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1)+P(A_2)P(B|A_2)}$$

$$=\frac{\frac{1}{2}\times 0.8}{\frac{1}{2}\times 0.8+\frac{1}{2}\times 0.1}=\frac{8}{9}\approx 0.89.$$

§1.5 相互独立随机事件，独立试验概型

一、相互独立随机事件

在例1.4.3中我们已经指出，当抽取产品是在有放回的情形，我们得到了 $P(B)=P(B|A)$ 这一事实，它说明事件 A 与事件 B 之间有某种"独立性"．对于具有这种性质的事件，我们较为直观的引入如下定义：

定义 1.5.1 设 $(\Omega,\mathscr{F},P)$ 为一概率空间，事件 $A\in\mathscr{F}$，$B\in\mathscr{F}$，且 $P(A)>0$，若 $P(B|A)=P(B)$，则称事件 B 随机独立于事件 A（或简称 B 独立于 A）．

性质 1.5.1 若事件 B 独立于 A，且 $P(A)>0$，$P(B)>0$，则事件 A 亦独立于 B．

事实上，由概率乘法公式(1.4.6)有

$$P(A)P(B|A)=P(B)P(A|B).$$

根据题设

$$P(B)=P(B|A),$$

得

$$P(A)=P(A|B).\ ▮$$

因此，当事件 A 与 B 满足 $P(A)>0$，$P(B)>0$ 时，它们之间的独立性具有相互对称性质，称它们相互独立．

性质 1.5.2 设 $(\Omega,\mathscr{F},P)$ 为一概率空间，$A\in\mathscr{F}$，$B\in\mathscr{F}$，且 $P(A)\cdot P(B)>0$，则事件 A 与 B 相互独立的充分必要条件为

$$P(A\cap B)=P(A)\cdot P(B). \tag{1.5.1}$$

证 必要性：由条件概率定义及事件A与B的相互独立性得

$$\frac{P(A\cap B)}{P(A)}=P(B|A)=P(B).$$

上述等式两边乘以$P(A)$得

$$P(A\cap B)=P(A)\cdot P(B).$$

充分性：因为

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{P(A)\cdot P(B)}{P(A)}=P(B),$$

即

$$P(B|A)=P(B). \blacksquare$$

(1.5.1)式对$P(A)=0$或$P(B)=0$亦成立.

事实上，如果$P(A)=0$，由$0\leqslant P(A\cap B)\leqslant P(A)=0$，可知$P(A\cap B)=0$，即(1.5.1)成立.

由上所述，因此我们引入如下两事件相互独立的一般定义.

定义 1.5.2 设$(\Omega,\mathscr{F},P)$是一概率空间，若事件A与B满足

$$P(A\cap B)=P(A)\cdot P(B),$$

则称事件A与B相互独立(这里不必规定$P(A)>0$或者$P(B)>0$).

性质 1.5.3 若事件A与B相互独立，则下列三对事件：$(A,\bar{B})$，$(\bar{A},B)$，$(\bar{A},\bar{B})$分别也相互独立.

证 由$A=A\cap\Omega=(A\cap B)\cup(A\cap\bar{B})$，得

$$P(A)=P(A\cap B)+P(A\cap\bar{B}),$$

由上式及A与B相互独立得

$$\begin{aligned}P(A\cap\bar{B})&=P(A)-P(A\cap B)=P(A)-P(A)\cdot P(B)\\&=P(A)[1-P(B)]=P(A)\cdot P(\bar{B}).\end{aligned}$$

故$(A,\bar{B})$相互独立. 由A，B的对称性，可见$(\bar{A},B)$也相互独立. 对于$(\bar{A},B)$重复应用上述证明方法，可得$(\bar{A},\bar{B})$亦相互

独立. ∎

例1.5.1 设甲、乙两射手独立地射击同一目标，他们击中目标的概率分别为0.9和0.8. 求在一次射击中，目标被击中的概率.

解 令 $A=\{$甲击中目标$\}$, $B=\{$乙击中目标$\}$.
依题意: $P(A)=0.9$; $P(B)=0.8$.

[方法一] 因为 A 与 B 相互独立，但是相容的，而"击中目标"这一事件 $C=A\cup B=\{$甲或乙击中目标$\}$，故

$$
\begin{aligned}
P(C)=P(A\cup B)&=P(A)+P(B)-P(A\cap B)\\
&=P(A)+P(B)-P(A)\cdot P(B)\\
&=0.9+0.8-0.9\times 0.8=0.98.
\end{aligned}
$$

[方法二]
$$
\begin{aligned}
P(C)&=1-P(\overline{C})=1-P(\overline{A\cup B})\\
&=1-P(\overline{A}\cap\overline{B})=1-P(\overline{A})\cdot P(\overline{B})\\
&=1-0.1\times 0.2=0.98.
\end{aligned}
$$

关于事件的独立性可推广到任意有穷多个事件的情形.

定义 1.5.3 设概率空间 $(\Omega,\mathscr{F},P)$ 中 n 个事件 $A_i\in\mathscr{F}(i=1,2,\cdots,n)$，若对任意 $s(1<s\leqslant n)$，及任意 $i_k; k=1,2,\cdots,s, 1\leqslant i_1<i_2<\cdots<i_s\leqslant n$，有

$$P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_s})=P(A_{i_1})\cdot P(A_{i_2})\cdots P(A_{i_s}), \tag{1.5.2}$$

则称 $A_1,A_2,\cdots,A_n$ 相互独立.

(1.5.2)式实际上包含有 2^n-1-n 个等式:

$P(A_{i_1}\cap A_{i_2})=P(A_{i_1})P(A_{i_2})$ 共 C_n^2 个等式;

$P(A_{i_1}\cap A_{i_2}\cap A_{i_3})=P(A_{i_1})P(A_{i_2})P(A_{i_3})$ 共 C_n^3 个等式;

……………………

$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2)\cdots P(A_n)$ 共 C_n^n 个等式. 故共有等式个数为

$$\sum_{i=2}^{n}C_n^i=\sum_{i=0}^{n}C_n^i-C_n^0-C_n^1=(1+1)^n-C_n^0-C_n^1=2^n-1-n.$$

从定义 1.5.3 看到，若 $A_1, A_2, \cdots, A_n$ 相互独立，则其中任意两个事件是独立的，但反过来却不一定正确.

例 1.5.2 设同时抛掷两个均匀的正四面体一次，每一四面体标有号码 1，2，3，4. 令

$A=\{$第一个四面体向下的一面出现偶数$\}$

$B=\{$第二个四面体向下的一面出现奇数$\}$

$C=\{$两个四面体向下的一面或者同时出现奇数，或者同时出现偶数$\}$

此时，样本空间为

$$\Omega=\begin{bmatrix}(1,1) & (1,2) & (1,3) & (1,4)\\(2,1) & (2,2) & (2,3) & (2,4)\\(3,1) & (3,2) & (3,3) & (3,4)\\(4,1) & (4,2) & (4,3) & (4,4)\end{bmatrix}.$$

由假设，利用古典概型我们可计算得诸事件概率如下：

$$P(A)=P(B)=P(C)=\frac{1}{2},$$

$$P(A\cap B)=P(A\cap C)=P(B\cap C)=\frac{4}{16}=\frac{1}{4}.$$

而

$$P(A\cap B\cap C)=0,$$

$$P(A)P(B)P(C)=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}.$$

故得

$$P(A\cap B)=P(A)P(B),$$

$$P(A\cap C)=P(A)P(C),$$

$$P(B\cap C)=P(B)P(C).$$

即任意两个事件都是独立的，但三个事件不是相互独立的，

因为

$$P(A\cap B\cap C)\neq P(A)P(B)P(C).$$

事件独立性的概念在概率论中是一个很重要的概念，不仅在理论上，而且在实用上都很重要．然而在实际应用中，因为它要计算 $P(A\cap B)$ 是否等于 $P(A)\cdot P(B)$，或 $P(B|A)$ 是否等于 $P(B)$，故要验证两事件的独立性比较困难，因此，往往根据问题的具体情况按独立性的实际意义来判定．当然，这是容易产生错误的．

下面计算几个有趣的例子．

例 1.5.3 如图 1.5.1 所示开关电路中，开关 a,b,c,d 开或关的概率均为 $\frac{1}{2}$，且是相互独立的，求

(1) 灯亮的概率；

(2) 已见灯亮，开关 a 与 b 同时关闭的概率．

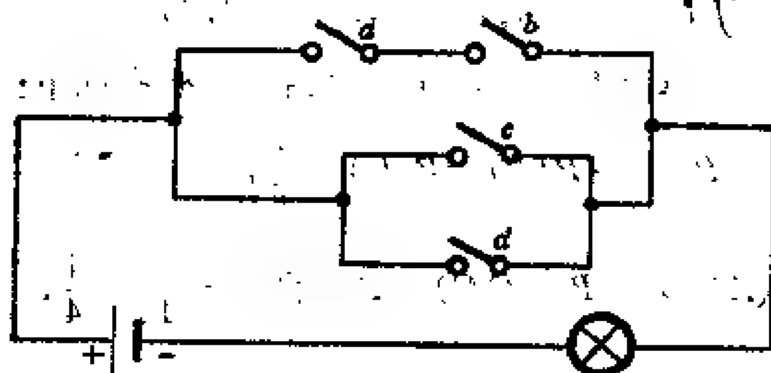


图 1.5.1

解 令 A,B,C,D 分别表示开关 a,b,c,d 关闭这些事件．

(1) 当 a、b 同时关闭或者 c、d 中有一个关闭时灯亮，即灯亮这事件为 $(A\cap B)\cup C\cup D$，由 (1.3.3) 式及相互独立性假设，可得灯亮的概率为

$$P\{(A\cap B)\cup C\cup D\}$$

$$=P(A\cap B)+P(C)+P(D)-P(A\cap B\cap C)$$

$$-P(A\cap B\cap D)-P(C\cap D)+P(A\cap B\cap C\cap D)$$

$$=P(A)P(B)+P(C)+P(D)-P(A)P(B)P(C)$$

$$-P(A)P(B)P(D)-P(C)P(D)+P(A)P(B)P(C)P(D)$$

$$=\left(\frac{1}{2}\right)^2+\frac{1}{2}+\frac{1}{2}-\left(\frac{1}{2}\right)^3-\left(\frac{1}{2}\right)^3-\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^4$$

$$=\frac{13}{16}.$$

(2) 由条件概率定义及题设，所求概率为

$$P\{A\cap B\mid(A\cap B)\cup C\cup D\}$$

$$=\frac{P\{(A\cap B)\cap[(A\cap B)\cup C\cup D]\}}{P[(A\cap B)\cup C\cup D]}$$

$$=\frac{P(A\cap B)}{P[(A\cap B)\cup C\cup D]}=\frac{P(A)P(B)}{\frac{13}{16}}$$

$$=\frac{\left(\frac{1}{2}\right)^2}{\frac{13}{16}}=\frac{4}{13}.$$

上面的计算过程中，我们应用了如下事实：由

$$A\cap B\subset(A\cap B)\cup C\cup D,$$

得

$$(A\cap B)\cap[(A\cap B)\cup C\cup D]=A\cap B.$$

例 1.5.4 设每支步枪射击飞机命中的概率为 $P=0.004$，求 250 支步枪同时独立地进行一次射击时，击中飞机的概率.

解 由题设知，250 支步枪全部都没有击中的概率为

$$(1-P)^{250}=(0.996)^{250}\approx0.37.$$

故所求的概率为

$$1-(1-P)^{250}\approx1-0.37=0.63.$$

由上结果看到，虽然每支步枪击中飞机的概率很小，但只要增加步枪的数目，击中飞机的概率可以大到接近 1 的程度.

例如：仍以上例题设为例，要以 0.99 的概率击中飞机，则所需

的步枪数 n 可由下式求得:

$$(1-P)^n=1-0.99,$$

即

$$(0.996)^n=0.01.$$

两边取对数得

$$n\lg(0.996)=\lg 0.01.$$

$$\therefore\ n=\frac{\lg 0.01}{\lg 0.996}\approx 1150.$$

即约需步枪 1150 支便能保证以 0.99 的概率击中飞机.

这个例子的思想在可靠性计算中有重要的应用. 下面介绍系统可靠度计算.

二、串联,并联系统的可靠度计算

可靠性理论随着工农业的飞快发展而得到迅速发展，特别是第二次世界大战以后,卫星、火箭、飞机、导航系统以及各种电子仪器等等,由于结构复杂,所含元件很多，人们又迫切需要提高各种设备系统的功效,这二者是有矛盾的,因为一个系统包含那么多的元件,只要某些元件发生故障，就有可能影响整个系统的工作. 因此,必须考虑解决这一矛盾. 可靠性理论就在这种情况下发展起来的. 目前,可靠性研究的内容大致可分为下面几个方面:

(1) 可靠性寿命试验——它是解决如何提高每一元件或部件的工作可靠性而进行对元件或部件的寿命试验，找出较好的配方或工艺过程,以求获得元件或部件的寿命的分布;

(2) 可靠性维护策略——为解决系统可靠性而研究维护(包括更新与检修)方法和理论;

(3) 系统可靠度计算——它主要是解决改良系统的结构，借以提高系统的可靠性,或者提出元件或系统可靠性指标.

下面只介绍系统可靠度计算的最简单内容，作为概率论应

用的例子.

所谓元件或系统的可靠度，通常是指在某一时间区间内元件或系统无故障工作(正常工作)的概率.

一个系统由许多元件按一定方式连接起来. 因此，系统的可靠度，依赖于每个元件的可靠度以及元件之间的连接形式. 下面只介绍如何通过元件可靠度来计算在元件的各种连接下的系统可靠度.

设系统由 n 个元件连接而成.

令 $A_i=\{$在时间区间$[0,t]$内第 i 个元件正常工作$\}$ $(i=1,2,\cdots,n)$，$A=\{$在时间区间$[0,t]$内系统正常工作$\}$. 并假定每一元件是否发生故障不影响其它元件的是否发生故障，即 $A_i(i=1,2,\cdots,n)$ 相互独立.

(1) 串联系统

若一个系统由 n 个元件按图 1.5.2 连接，称为串联系统.

它的特点是当其中一个元件发生故障时整个系统就发生故障，亦即系统能正常工作的充分必要条件为 n 个元件同时正常工作.



图 1.5.2

因此有

$$P(A)=P(A_1\cap A_2\cap\cdots\cap A_n).\tag{1.5.3}$$

由独立性假设，(1.5.3)式变为

$$P(A)=P(A_1)P(A_2)\cdots P(A_n).\tag{1.5.4}$$

从(1.5.4)式看到，当 n 越大(即元件越多)，系统可靠度越小，如设 $P(A_i)=\frac{1}{2}(i=1,2,\cdots,n)$，则

$$P(A)=\left(\frac{1}{2}\right)^n \longrightarrow 0 \quad (n\to\infty).$$

因此，对串联系统要提高可靠性，必须要求元件数量越小越好．当然，在实际上，将因其它指标的要求而不能无限制的减少元件．

(2) 并联系统

一个系统若有 n 个元件按图 1.5.3 连接，称为并联系统．

它的特点是当而且仅当 n 个元件全部发生故障时系统发生故障，即当 n 个元件中有一个以上的元件正常工作时，系统就能正常工作．在这种并联系统情况下，系统在时间区间 $[0,t]$ 正常工作的概率

$$P(A)=P\left(\bigcup_{i=1}^{n} A_i\right).$$

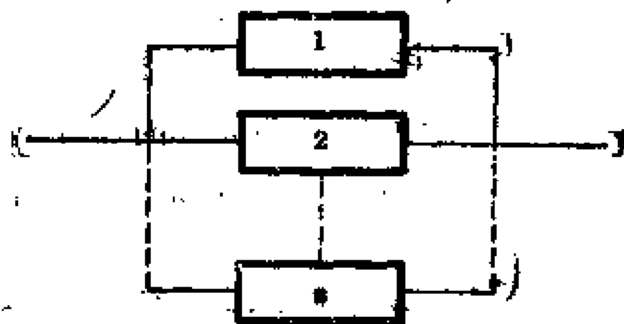


图 1.5.3

注意到各元件是否正常工作的独立性假设，上式可通过如下运算得到，

$$\begin{aligned} P(A)&=1-P(\bar{A})=1-P\left(\overline{\bigcup_{i=1}^{n} A_i}\right) \\ &=1-P\left(\bigcap_{i=1}^{n} \bar{A}_i\right)=1-P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_n) \\ &=1-\prod_{i=1}^{n}[1-P(A_i)]. \end{aligned} \tag{1.5.5}$$

从(1.5.5)式看到，并联系统可靠度将因元件的增加而增大，例如 $P(A_i)=\frac{1}{2}$，由(1.5.5)式得

$$P(A)=1-\prod_{i=1}^{n}\left(1-\frac{1}{2}\right)=1-\left(\frac{1}{2}\right)^{n}\to 1 \qquad (n\to\infty).$$

三、独立试验概型

独立试验概型在概率论理论和应用方面起着十分重要的作用.我们在 §1.1 的一开始就提到过，随机现象的统计规律性只有在相同条件下进行大量的重复试验或观察才呈现出来. 这些重复试验的特点是

i) 每次条件都一样，且可能的结果为有限个；

ii) 各次试验的结果不互相影响，或者称为相互独立.

特点 i) 意味着每次试验所对应的概率空间都是一样的，设为 $(\Omega,\mathscr{F},P)$，其基本事件数为有限.

特点 ii) 表示，"第 1 次试验出现的事件为 A_{i_1}，第 2 次试验出现的事件为 A_{i_2}，…，第 n 次试验出现的事件为 A_{i_n}" 这一联合事件的概率为

$$P(A_{i_1})P(A_{i_2})\cdots P(A_{i_n}).$$

满足i)，ii)的 n 次重复试验称为n次独立试验概型. 特别，当每次试验的基本事件只有两种，即只有两个事件 A 及 $\bar{A}$，且 $P(A)=p$，$P(\bar{A})=q=1-p$ 时，称为贝努里概型. 所以一般情况就称为 n 次贝努里概型.

定理 1.5.1 对于贝努里概型，事件 A 在 n 次试验中出现 k 次的概率为

$$P_n(k)=C_n^k p^k q^{n-k} \qquad (0\leqslant k\leqslant n). \qquad (1.5.6)$$

并且

$$\sum_{k=0}^{n}P_n(k)=1. \qquad (1.5.7)$$

证 由贝努里概型知，事件 A 在指定的 k 次试验中发生，而在其余 $n-k$ 次试验中不发生的概率为

$$p^k(1-p)^{n-k}=p^k q^{n-k}.$$

(例如"在前 k 次试验发生事件 A 而其余 $n-k$ 次试验不发生事件 A"的概率为 $\underbrace{pp\cdots p}_{k\text{个}}\ \underbrace{qq\cdots q}_{n-k\text{个}}=p^k q^{n-k}$)

由于事件 A 的发生可以有各种排列顺序，由排列组合理论知它共有 C_n^k 种，而 C_n^k 种排列所对应的 C_n^k 个事件(即"事件 A 在 n 次试验中出现 k 次"这一事件)是互不相容的。按概率加法公式得到

$$P_n(k)=C_n^k p^k q^{n-k},$$

并且有

$$\sum_{k=0}^{n}P_n(k)=\sum_{k=0}^{n}C_n^k p^k q^{n-k}=(p+q)^n=1. \quad ▮$$

下面，我们举几个例子，说明如何应用定理 1.5.1 去计算概率.

例 1.5.5 设有八门火炮独立地同时向一目标各射击一发炮弹，若有不少于 2 发炮弹命中目标时，目标算作被击毁，如果每门炮命中目标概率为 0.6，求击毁目标的概率 p 是多少？

解 依题意，令 A 表示每一门火炮命中目标这一事件，则 $P(A)=0.6$. 这样，例 1.5.5 可看作 $n=8$ 的贝努里概型问题，按定理 1.5.1 得所求概率为

$$p=\sum_{k=2}^{8}P_8(k)=1-P_8(0)-P_8(1)$$

$$=1-C_8^0(0.6)^0(0.4)^8-C_8^1(0.6)^1(0.4)^7=0.991.$$

例 1.5.6 对某种药物的疗效进行研究，假定这药物对某种疾病的治愈率 $P_0=0.8$，现在 10 个患此病的病人同时服用此药，求其中至少有 6 个病人治愈的概率 p.

解 假定病人服用该药后或者治愈(事件 A)，或者没有治愈(事件 $\bar{A}$)，按题设

$$P(A)=0.8, P(\bar{A})=0.2.$$

因此可用 $n=10$ 的贝努里概型去求其概率，得

$$p=\sum_{k\geq 6}^{10} P_{10}(k)=P_{10}(6)+P_{10}(7)+P_{10}(8)+P_{10}(9)+P_{10}(10)$$

$$=C_{10}^{6}(0.8)^{6}(0.2)^{4}+C_{10}^{7}(0.8)^{7}(0.2)^{3}+C_{10}^{8}(0.8)^{8}(0.2)^{2}+C_{10}^{9}(0.8)^{9}(0.2)^{1}+C_{10}^{10}(0.8)^{10}(0.2)^{0}\approx 0.97.$$

这结果表示在 100 次药物服用试验中，大约有 97 次使得 10 人中至少有 6 人被治愈．换句话说，如果治愈率确为 0.8，则在 10 个病人服用此药物后治愈人数少于 6 人这一事件是很少出现的(概率为 0.03)．利用这一结果，若在一次实际服用此药试验中，10 个病人治愈不到 6 人，我们就要对治愈率为 0.8 表示怀疑．它说明实际上治愈率小于 0.8．

例 1.5.7 设某种数字传输制以每秒 512×10^{3} 个 0 或 1 的序列传送消息(即每秒传送 512×10^{3} 个 0 或 1)．由于各种干扰，传送过程中会产生将 0 误为 1 或将 1 误为 0 的情况，这两种情况均称为"误码"．设误码的概率 p_0(误码率)为 10^{-7}，求在 10 秒内出现 1 个误码的概率 p 是多少？

解 这问题可看作 $n=512\times 10^{3}\times 10$ 的贝努里试验，因此所求概率 $p=C_{512\times 10^{4}}^{1}10^{-7}(1-10^{-7})^{512\times 10^{4}-1}$．

显然，要直接计算上述概率是很困难的，需要用下面近似公式

$$P_n(k)\approx\frac{\lambda^k}{k!}e^{-\lambda} \text{①}, \tag{1.5.8}$$

① 公式的证明将在第二章中给出。

其中 $\lambda = np_0$. 我们应用(1.5.8)式的近似公式计算 例 1.5.7 的概率. 此时

$$n=512\times 10^3\times 10,\ p_0=10^{-7},$$

$$\lambda=np_0=512\times 10^4\times 10^{-7}=0.512.$$

故所求概率为

$$p\approx\frac{\lambda}{1!}e^{-\lambda}$$

$$=0.512\times e^{-0.512}$$

$$\approx 0.512\times 0.6\approx 0.3.$$

例 1.5.8 (巴拿赫问题) 某人的口袋中经常装有两盒火柴,每盒 n 枝. 使用时,自此两盒中等可能地任选一盒,然后从其中抽取一枝. 某次,此人发现首次用完一盒,问此时另盒中恰有 r ($r=0, 1, \cdots, n$) 枝火柴的概率是多少?

解 这时此人共使用了 $2n-r$ 枝火柴,欲有一盒中剩 r 枝,必须且只须在此盒中使用 n 枝,另一盒使用了 $n-r$ 枝. 由于每次使用中使用到每盒的概率为 $\frac{1}{2}$, 故在 $2n-r$ 次使用中使用到某盒 n 次和使用另一盒 $n-r$ 次, 故所求概率为

$$p=2C_{2n-r}^{n}\left(\frac{1}{2}\right)^{n}\cdot\left(\frac{1}{2}\right)^{n-r}\cdot\left(\frac{1}{2}\right)=2C_{2n-r}^{n}\left(\frac{1}{2}\right)^{2n-r}$$

(上式右边之所以有乘数因子 2,是考虑到两火柴盒的对称性).

当 $r=0$ 时,

$$p=2C_{2n}^{n}\left(\frac{1}{2}\right)^{2n}.$$

当 $r=n$ 时,

$$p=\left(\frac{1}{2}\right)^{n-1}.$$

从上式看到,当 $n\to\infty$ 时, $p\to 0$. 它表示, 当每盒的火柴枝数

较大，而"发现首次用完一盒，而另一盒中一枝火柴都没有用上"这一可能性是很小的.

独立试验概型的精确数学描述，必须引入乘积空间概念：每次试验联系于一个概率空间 $(\Omega, \mathscr{F}, P)$，连续两次试验得一点偶 (ω_1, ω_2)，其中 $\omega_1 \in \Omega$，$\omega_2 \in \Omega$；连续 n 次试验得一 n 维空间的点 $(\omega_1, \omega_2, \cdots, \omega_n)$ 其中 $\omega_i \in \Omega$ $(i=1, 2, \cdots, n)$. 因此，连续 n 次试验的样本空间是

$$\Omega_n = \{(\omega_1, \omega_2, \cdots, \omega_n);\ \omega_i \in \Omega \ \ (i=1, 2, \cdots, n)\},$$

记它为

$$\Omega_n = \underbrace{\Omega \times \Omega \times \cdots \times \Omega}_{n\text{个}} = \Omega^n;$$

并称之为 n 维乘积空间. 例如如 $\Omega = R$，二维实数乘积空间就是平面，三维实数乘积空间就是三维欧氏空间，等等. 又如连续随机抛掷硬币，则二次连续抛掷硬币的样本空间为 $\Omega_2 = \Omega^2 = \{(\text{正},\text{正}), (\text{正},\text{背}), (\text{背},\text{正}), (\text{背},\text{背})\}$，三次连续抛掷硬币的样本空间为

$$\Omega_3 = \Omega^3 = \{(\text{正},\text{正},\text{正}), (\text{正},\text{正},\text{背}), \cdots, (\text{背},\text{背},\text{背})\}.$$

样本空间 Ω_n 中一般用下法定义 σ-代数 $\mathscr{F}_n$：如下形式的集合

$$A = \{(\omega_1, \omega_2, \cdots, \omega_n);\ \omega_i \in A_i,\ A_i \in \mathscr{F},\ i=1, 2, \cdots, n\}$$

称为矩形集. 如平面上的矩形 $A = A_1 \times A_2$，其中 $A_1 = \{x;\ a \leqslant x \leqslant b\}$，$A_2 = \{y;\ c \leqslant y \leqslant d\}$，图 1.5.4 阴影部分表示 $A = A_1 \times A_2$.

$\mathscr{F}_n$ 定义为包含全体矩形集的最小 σ-代数.

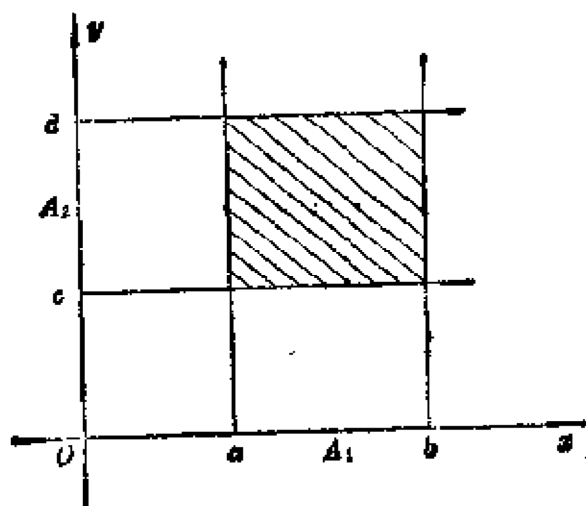


图 1.5.4

$(\Omega_n, \mathscr{F}_n)$ 上可以定义概率 P_n，对于 $A=A_1\times A_2\times\cdots\times A_n$，其对应的概率为 $P_n(A_1\times A_2\times\cdots\times A_n)$. n 次试验对应一个概率空间 $(\Omega_n, \mathscr{F}_n, P_n)$. 若对任一 $A\in\mathscr{F}_n$，其中

$$A=A_1\times A_2\times\cdots\times A_n,$$

$$P_n(A_1\times A_2\times\cdots\times A_n)=P(A_1)P(A_2)\cdots P(A_n),$$

则称 n 次试验是独立的.

习　题

1. 设 $\Omega=\{1,2,3,4,5,6,7,8,9,10\}$, $A=\{2,3,4\}$, $B=\{3,4,5\}$, $C=\{5,6,7\}$，求下列事件：

i) $\overline{\bar{A}\cap\bar{B}}$;　　　ii) $\overline{A\cap(\overline{B\cap C})}$.

2. 利用事件的运算和关系证明下列各式：

i) $\bar{\bar{A}}=A$;

ii) $(A-A\cap B)\cup B=A\cup B=\overline{(\bar{A}\cap\bar{B})}$;

iii) $(A\cup B)-B=A-(A\cap B)=A\cap\bar{B}$;

iv) $(A\cup B)-(A\cap B)=(A\cap B)\cup(\bar{A}\cap B)$.

3. 证明下列恒等式：

i) $C_n^r=C_{n-1}^r+C_{n-1}^{r-1}$;

ii) $C_n^{r+1}=C_{n-1}^r+C_{n-2}^r+\cdots+C_r^r$.

4. 试证明：

i) $C_{n-m}^r\cdot C_m^0+C_{n-m}^{r-1}\cdot C_m^1+\cdots+C_{n-m}^0\cdot C_m^r=C_n^r$;

ii) $\dfrac{C_m^k\cdot C_{n-m}^{r-k}}{C_n^r}=\dfrac{C_r^k\cdot C_{n-r}^{m-k}}{C_n^m}$,

并说明其概率意义.

5. 设 n 个人排成一行，甲与乙是其中的两个人，求这 n 个人的任意排列中，甲与乙之间恰有 r 个人的概率. 如果 n 个人围成一圆圈，试证明甲与乙之间恰有 r 个人的概率与 r 无关，都是 $\dfrac{1}{n-1}$（在圆圈排列时，仅考虑从甲到乙的顺时针方向）.

6. 在整数 0 至 9 中任取 4 个，能排成一个四位偶数的概率是多少？

7. 口袋内放有 2 个伍分、3 个贰分、5 个壹分钱的硬币，任取其中 5 个，

求总值超过一角钱的概率.

8. 箱中盛有 α 个白球和 β 个黑球，从其中任意地接连取出 $k+1(k+1\leqslant\alpha+\beta)$ 个球，如每球取出后不放回，试求最后取出的是白球的概率.

9. 一架电梯开始时有 6 位乘客并等可能地停于 10 层楼的每一层，求下列事件的概率：

i）某一层有两位乘客离开；

ii）没有两位及两位以上乘客在同一层离开；

iii）恰有两位乘客在同一层离开；

iv）至少有两位乘客在同一层离开.

（假定乘客离开的各种可能排列具有相同的概率.）

10. 一列火车共有 n 节车厢，有 $k(k\geqslant n)$ 个旅客上火车并随意地选择车厢. 求每一节车厢内至少有一个旅客的概率.

11. 设 $P(A)=x, P(B)=y$ 且 $P(A\cap B)=z$. 用 x, y, z 表示下列事件的概率：

i）$P(\bar{A}\cup B)$；　　　　ii）$P(\bar{A}\cap B)$；

iii）$P(\bar{A}\cup\bar{B})$；　　　　iv）$P(\bar{A}\cap\bar{B})$.

12. 证明 $P[(A\cap\bar{B})\cup(B\cap\bar{A})]=P(\bar{A})+P(\bar{B})-2P(\bar{A}\cap\bar{B})$，并解释此结果的概率意义.

13. 设事件 A、B、C 满足：

$$P(A)=P(B)=P(C)=\frac{1}{4},\qquad P(AB)=P(CB)=0,$$

$$P(AC)=\frac{1}{8}.$$

试求事件 A、B、C 至少有一个发生的概率.

14. 将线段 $(0, a)$ 任意折成三折，试求此三折线段能构成三角形的概率.

15. 甲、乙两艘轮船驶向一个不能同时停泊两艘轮船的码头停泊，它们在一昼夜内到达的时刻是等可能的. 如果甲船的停泊时间是一小时，乙船的停泊时间是二小时，求它们中的任何一艘都不需等待码头空出的概率.

16. 试证：如果 $P(A|B)>P(A)$，则 $P(B|A)>P(B)$.

17. 一批产品共 100 件，对其进行抽样检查，整批产品看作不合格的规定是：在被检查的 5 件产品中只要有一件是废品. 如果在该批产品中有 5% 是不合格品，试问该批产品被认为不合格的概率是多少？

18. 全部产品中 4% 是废品，而合格品中的 75% 为一级品，求任选一个

产品为一级的概率.

19. 当 $P(A)=a, P(B)=b$ 时，证明：

$$P(A|B) \geqslant \frac{a+b-1}{b}.$$

20. 进行摩托车竞赛. 在地段甲、乙间布设了三个故障，在每一故障前停车的概率为 0.1. 从乙地到终点丙地竞赛者不停车的概率为 0.7，求在地段甲、丙间竞赛者不停车的概率.

21. 卜里耶概型——设口袋里装有 b 个黑球，r 个红球，任意取出一个，然后放回并再放入 c 个与取出的颜色相同的球，再向袋里取出一球，问：

i) 最初取出的球是黑的，第二次取出的也是黑色的概率；

ii) 如将上述手续进行 n 次，取出的正好是 n_1 个黑球，n_2 个红球（$n_1+n_2=n$）的概率；

iii) 用归纳法证明：任何一次取得黑球的概率都是 $\frac{b}{b+r}$；任何一次取得红球的概率都是 $\frac{r}{b+r}$；

iv) 用归纳法证明：第 m 次与第 n 次（$m<n$）取出都是黑球的概率是 $\frac{b(b+c)}{(b+r)(b+r+c)}$.

22. 利用概率论的想法证明恒等式（其中 $A>a$ 均为正整数）：

$$1+\frac{A-a}{A-1}+\frac{(A-a)(A-a-1)}{(A-1)(A-2)}+\cdots+\frac{(A-a)\cdots 3\cdot 2\cdot 1}{(A-1)\cdots(a+1)a}=\frac{A}{a}.$$

23. 两批相同的产品各有 12 件和 10 件，在每批产品中有一件废品. 今任意地从第一批中抽出一件混入第二批中，然后再从第二批中抽出一件，求从第二批产品中抽出的是废品的概率.

24. 在一盒子中装有 15 个乒乓球，其中有 9 个新球. 在第一次比赛时任意取出三个球，比赛后仍放回原盒中；在第二次比赛时同样任意取出三个球，求第二次取出的三个球均为新球的概率.

25. 为了传递消息，采用电报系统发出"点"和"划"的信号. 根据统计，干扰的情况是：传送"点"时平均有 $\frac{2}{5}$ 失真，而传送"划"时有 $\frac{1}{3}$ 失真. 已知在传送的信号中，"点"与"划"之比为 5:3，求在接收的信号中，"点"与"划"恰好是发出信号的"点"与"划"的概率.

26. 某工厂有三部制螺钉的机器 A、B、C，它们的产品分别占全部产品

的 25%，35%，40%，并且它们的废品率分别是 5%、4%、2%．今从全部产品中任取一个，并发现它是废品．问它是 A、B、C 制造的概率各为多少？

27．某仪器有三个灯泡，烧坏第一、第二、第三个灯泡的概率相应地为 0.1、0.2 及 0.3，并且相互独立．当烧坏一个灯泡时，仪器发生故障的概率为 0.25，当烧坏两个灯泡时为 0.6，而当烧坏三个时为 0.9．求仪器发生故障的概率．

28．已知 $P(B|A)=P(B|\bar{A})$，证明事件 A、B 相互独立．

29．甲、乙比赛射击，每进行一次，胜者得一分．在一次射击中，甲"胜"的概率为 α，乙"胜"的概率为 β．设 $\alpha>\beta(\alpha+\beta=1)$，且独立地进行比赛到有一人超过对方 2 分就停止，多得 2 分者胜．求甲、乙获胜的概率．

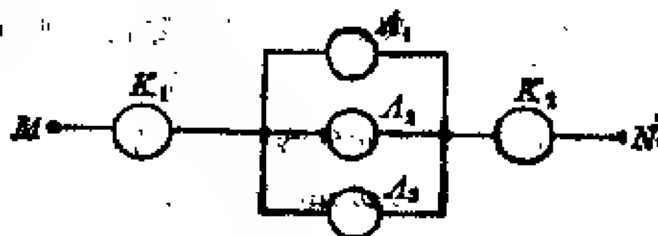
30．（小概率事件）设随机试验中某一事件 A 出现的概率 $\varepsilon>0$，求在三次独立试验中 A 出现的概率．并证明：不断独立地重复做此试验时，A 迟早会出现的概率为 1，不论 $\varepsilon>0$ 如何小．

31．进行四次独立的试验，在每一次试验中 A 出现的概率为 0.3．如果 A 不出现，则 B 也不出现；如果 A 出现一次，则 B 出现的概率为 0.6；如果 A 出现不少于二次，则 B 出现的概率为 1．试求 B 出现的概率．

32．在四次独立试验中事件 A 至少出现一次的概率为 0.59，试问在一次试验中 A 出现的概率是多少．

33．如果在一次测量中得到正值误差的概率等于 $\frac{2}{3}$，得负值误差的概率为 $\frac{1}{3}$，求在四次测量中具有最大概率的负值误差次数及其相对应的概率．

34．在 M 与 N 两点之间的电路按下图构成：



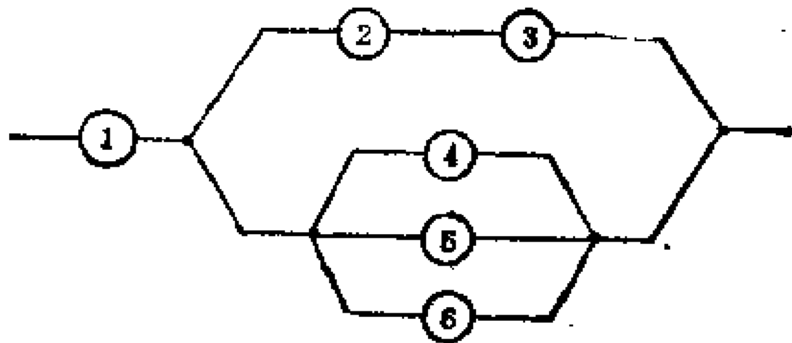
（第 34 题图）

在时间 T 内电路中不同元件发生故障是独立事件，其概率为：

元件：	K_1	K_2	A_1	A_2	A_3
概率：	0.6	0.5	0.4	0.7	0.9

求在指定的时间中由于：i）K_1或K_2发生故障而断电的概率；ii）A_1、A_2、A_3同时发生故障而断电的概率；iii）K_1或K_2或A_1、A_2、A_3同时发生故障而断电的概率.

35. 设有一部件结构示意图如下：



（第35题图）

其中6个零件是独立工作的，④、⑤、⑥三个中，有两个是备用件（当正在工作的那一个失效时，其中另一个立即补充上去）．求部件正常工作的概率．（设$P\{i$正常工作$\}=p_i(i=1,2,3)$，$P\{$④正常工作$\}=P\{$⑤正常工作$\}=P\{$⑥正常工作$\}=p$）.

36. 在每一次试验中，事件A出现的概率为p，试问在n次独立试验中A出现偶数次的概率是多少？

37 在间隔时间t内向电话总机呼叫k次的概率为$P_t(k)$，若在任意两个相邻的间隔时间内呼叫次数是相互独立的，求在间隔时间$2t$内呼叫s次的概率$P_{2t}(s)$.

38. 已知自动织布机在Δt这段时间内因故障而停机的概率为$a\cdot\Delta t+o(\Delta t)$（$a$是常数），并设机器在不重迭时间内停机的各个事件是彼此独立的．假定在时刻t_0机器在工作着，试求此机器在由时刻t_0到t_0+t这段时间内不停止工作的概率$P(t)$（设$P(t)$与初始时刻t_0无关）.

第二章　随机变数及其分布函数

§ 2.1　随机变数的直观意义与定义

在第一章我们曾经指出过，观察一个随机现象，其样本点(即出现的结果)可以是数量性质的，也可以是非数量性质的．前者如某电话总机在时间区间$(0, T]$内收到的呼叫次数是0次，1次，…；后者如某次射击的"中的"与"不中的"．或者是一个随机试验的结果，可以是"成功"，也可以是"失败"．

概率论与数理统计是这样的一门学科，它从数量的侧面来研究随机现象的统计规律性，建立起一系列的公式和定理，借以更好地描述、处理与解决各种和随机现象有关的理论和应用问题．

为便于处理起见，让我们只讨论具有数量性质的随机试验或观察．假设它可能出现的全部结果都能数量化；而且，我们有一个法则，使得按照这个法则，每个结果，只对应着一个数(或每个结果，都只对应着一个确定维数的向量——这种情形将在§2.2中考虑)．若是这样，则对这一个随机试验或观察的研究，便可以用一个具有"随机"取值的变数(即所谓随机变数)来处理与研究．它的定义域中的点是实验出现的"结果"（或称"样本点"）——称为"基本事件"，值域中的点便是与此"结果"相联系的用上述某种法则所对应的唯一确定的"数"；并且，由于它取值的"随机性"，我们还应要求"它取值于某一给定的区间"这种事件为随机事件．让我们先来看可以将"样本点"与"数"对应起来的一些例子．

例2.1.1　对于某电话总机在时间区间$(0, T]$内收到的呼叫

次数，我们可以指定数 0,1,2,…分别与没有呼叫，一次呼叫，二次呼叫，…相对应．这样我们就建立了 $\Omega=\{k$ 次呼叫 ω_k；$k=0,1,2,\cdots\}$ 与 0 和全体正整数的对应关系．

例 2.1.2 考虑在相同条件下某车床加工的另件长度与规定的长度的偏差 ω_{cm}，由于通常是可以知道其偏差的范围的；故可以假定其偏差的绝对值小于某正数 a．若是这样，则对于每一个偏差 $\omega_{cm}\in[-a, a]$，可指定 $y=\omega_{cm}\in[-a,a]$ 与之相对应，这样我们就建立了 $\Omega=\{\omega_{cm}:\omega_{cm}\in[a, -a]\}$ 与区间 $[-a, a]$ 之间的对应关系。

对于非数量性质的随机现象也可以这样做．

例 2.1.3 在向上抛掷均匀硬币的情形，样本空间可取为 $\Omega=\{\omega_1,\omega_2\}$，其中 ω_1,ω_2 分别代表硬币落地时正面朝上或背面朝上的两个样本点；我们可以指定数 1 与 ω_1 对应，数 0 与 ω_2 对应．这样就建立了 $\Omega=\{\omega_1,\omega_2\}$ 与两个数 1 和 0 的对应关系．

上述样本空间与实数空间的对应关系，常用希腊字母 ξ，η，ζ，…表示，它们取的值常用英文小写字母 x,y,z，…表示．例如

在例 2.1.1 中有 $\xi(\omega_0)=0,\xi(\omega_1)=1,\xi(\omega_2)=2,\cdots$；

在例 2.1.2 中有 $\eta(\omega_{cm})=y=\omega_{cm},\omega_{cm}\in[-a,a]$，

$$y\in[-a,a];$$

在例 2.1.3 中有

$$\zeta(\omega_1)=1,\zeta(\omega_2)=0.$$

上面的例子，给出了"样本点"与"数"之间建立对应关系的情形，亦即，对于每一个样本点 ω，可有一个数 $\xi(\omega)$ 与之对应．但是，为了能确定 $\xi(\omega)$ 取某些值的概率，亦即为了能刻划 $\xi(\omega)$ 取值的"随机性"，还需要对 $\xi(\omega)$ 加以一定的限制与规定．为此，引入如下的随机变数的数学定义：

定义 2.1.1 设 $(\Omega,\mathscr{F},P)$ 是一个概率空间，对于 $\omega\in\Omega$，$\xi(\omega)$

是一个取实值的单值函数；若对于任一实数 x，$\{\omega:\xi(\omega)<x\}$① 是一随机事件，亦即$\{\omega:\xi(\omega)<x\}\in\mathscr{F}$，则称 $\xi(\omega)$ 为随机变数.

从定义 2.1.1 看到，随机变数 $\xi(\omega)$ 总是联系着一个概率空间$(\Omega,\mathscr{F},P)$. 为书写简便，今后没有特殊必要，我们不必每次都写出概率空间$(\Omega,\mathscr{F},P)$；并且将随机变数 $\xi(\omega)$写为 ξ，省去 ω；而把 $\{\omega:\xi(\omega)<x\}$记为 $\{\xi<x\}$ 等等. 另一方面，由于我们要求$\{\omega:\xi(\omega)<x\}\in\mathscr{F}$，因此 $P\{\xi<x\}$总是有意义的.

随机变数 ξ 是一个定义在抽象空间上的实函数，而不是象数学分析中通常定义的函数那样.

在同一概率空间上可以定义许多随机变数，只要它满足定义 2.1.1 即可.

定义在$(\Omega,\mathscr{F},P)$上的单值函数 ξ 是否为随机变数，要视$(\Omega,\mathscr{F},P)$中的 σ-代数的结构如何而定.（亦即$\mathscr{F}$能否包含一切形如$\{\xi<x\}$的集合，其中 x 为任意实数.）

由 σ-代数$\mathscr{F}$的性质及随机变数的定义可得

$$\{\xi=x\}=\bigcap_{k=1}^{\infty}\left\{\xi<x+\frac{1}{k}\right\}-\{\xi<x\}\in\mathscr{F};$$

$$\{\xi\leqslant x\}=\{\xi<x\}\cup\{\xi=x\}\in\mathscr{F};$$

$$\{\xi>x\}=\Omega-\{\xi\leqslant x\}\in\mathscr{F}.$$

类似地，还可证$\{\xi\geqslant x\}$，$\{x_1<\xi<x_2\}$，$\{x_1\leqslant\xi\leqslant x_2\}$等等都属于$\mathscr{F}$，故都有确定的概率. 因此随机变数可以采取别的等价定义. 例如，可作下面等价定义：若对一切实数 x，有$\{\xi(\omega)\leqslant x\}\in\mathscr{F}$，则称 $\xi(\omega)$为随机变数.

从例 2.1.1，例 2.1.2 及例 2.1.3 可见，随机变数取值的情形可以很不相同，有的只取有限个数值（如例 2.1.3，只取 0 与 1 为值），或可列无穷多个值（如例 2.1.1，取值为 0，1，2，…）；这一类随机变数，我们称之为离散型随机变数. 另一类随机变数则是可以取值于某一区间中的任一数（如例 2.1.2，取值于$[-a,a]$），这种

① $\{\omega:\xi(\omega)=x\}$表示满足 $\xi(\omega)=x$的 ω 全体，$\{\omega:\xi(\omega)<x\}$表示满足 $\xi(\omega)<x$ 的 ω 全体，余类推.

随机变数我们称之为非离散型随机变数。

下面我们就分开情形讨论.

一、离散型随机变数与分布列

定义 2.1.2 设 ξ 为离散型随机变数，亦即 ξ 的一切可能值为 $x_1, x_2, \cdots, x_n, \cdots$ 记 $p_n = P(\xi = x_n)\ (n=1,2,\cdots)$，称 $p_1, p_2, \cdots, p_n, \cdots$ 为 ξ 的分布列，亦称为 ξ 的概率函数①.

由上述可知，若 ξ 为随机变数，则 p_n 有意义.

对于离散型随机变数，把它列出下表更为直观②：

ξ	x_1	x_2	$\cdots$	x_n	$\cdots$
$P(\xi=x_n)$	p_1	p_2	$\cdots$	p_n	$\cdots$

显然，其中 p_n 满足下列两个条件：

$$\left.\begin{array}{l} 1)\ \ p_n \geqslant 0\ (n=1,2,\cdots), \\ 2)\ \ \sum_{n=1}^{\infty} p_n = 1. \end{array}\right\} \tag{2.1.1}$$

利用微积分知识，容易证明：满足(2.1.1)的序列 $\{p_n\}$，必存在一个 n_0，使得

$$p_{n_0} = \max_n (p_n);$$

由于

$$p(\xi = x_{n_0}) = p_{n_0},$$

故自然称 x_{n_0} 为随机变数 ξ 的最大可能值.

下面举出一些常用的离散型随机变数的例子。

例 2.1.4 (退化分布或单点分布)

ξ	c
$P(\xi=c)$	1

分布列如上表所示的随机变数 ξ 显然概率为 1 地取值 c. 称

① 若 ξ 只有有限个可能值：$x_1, x_2, \cdots, x_n$；则自然称 $p_1, p_2, \cdots, p_n$ 为 ξ 的分布列.

② 分布列由表所唯一确定，故表不妨就称为分布列.

之为退化分布或单点分布

例 2.1.5 (两点分布)设随机变量 ξ 的分布列如下表:

ξ	0	1
$p(\xi=x)$	$1-p$	p

则称 ξ 服从两点分布(其中设 $0\leqslant p\leqslant 1$).

显然上表满足(2.1.1)式.

对于离散型随机变量的分布列,用分布在一条直线上的,集中了不同质量的点(简称质点)来描绘,似乎更为直观. 例如,对于两点分布,可用两个质点表示:用在 $x=0$ 点集中了质量 $1-p$,在 $x=1$ 点集中了质量 p 的(概率)质量分布图来描绘,如图 2.1.1.



(设 $p=0.2$)

图 2.1.1 两点分布的(概率)质量分布图

图 2.1.1 清楚地表明了:对于 $p=0.2$ 的情形,位于 $x=0$ 处之质点,其质量大(即取 $x=0$ 值的概率大),位于 $x=1$ 处的质点,其质量小(即取 $x=1$ 值的概率小).

分布列也可以用概率函数 $p(x_n)$ 图来描绘. 如图 2.1.2. 它同样清楚地说明,取 0 值与取 1 值概率的大小.

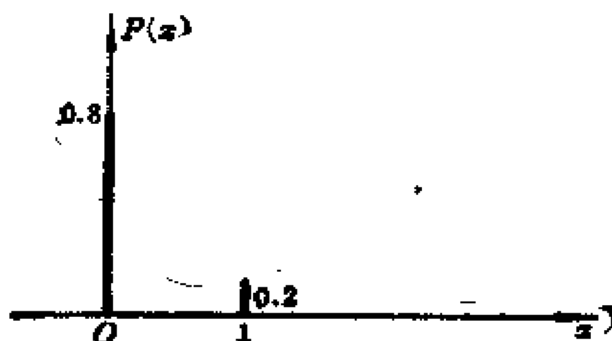


图 2.1.2 两点分布的概率函数 $p(x_n)$ 图(设 $p=0.2$)

二点分布可以作为描绘射手射击“中的”(此时，令随机变数 ξ 取值“1”)与“不中的”(此时，令随机变数 ξ 取值“0”)的概率分布情况的一个数学模型，或作为随机抛掷硬币落地时出现“正面”(相当于“中的”)与“背面”(相当于“不中的”)的概率分布情况的数学模型．当然也可作为从一批产品中任意抽取一件得到的是“正品”或“废品”(非正品)的模型．

例 2.1.6(二项分布)　若随机变数 ξ 由下表给出：

ξ	0	1	…	k	…	n
$P\{\xi=k\}$	$(1-p)^n$	$C_n^1p(1-p)^{n-1}$	…	$C_n^kp^k(1-p)^{n-k}$	…	p^n

(其中　$0\leqslant p\leqslant 1$，　n 为非负整数．)

则称 ξ 服从二项分布 $b(n,p)$．

易知上表满足(2.1.1)式．事实上，$P\{\xi=k\}\geqslant 0$ 是显然的；再由二项展开式知

$$\sum_{k=0}^{n}P(\xi=k)=\sum_{k=0}^{n}C_n^kp^k(1-p)^{n-k}=(p+q)^n=1,$$

其中 $q=1-p$．

由表可见，随机变数 ξ 取值 k 的概率 $P(\xi=k)=C_n^kp^kq^{n-k}$ 恰好是 $(p+q)^n$ 这二项展式的第 $k+1$ 项，这就是二项分布名称的由来．

图 2.1.3 与图 2.1.4 分别给出了二项分布 $b(3,0.3)$ 的(概率)质量分布图与概率函数图．其中

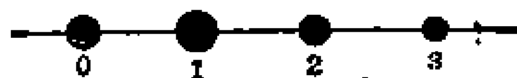


图 2.1.3　$b(3,0.3)$ 的(概率)质量分布图

$$P(\xi=0)=0.3430;$$

$$P(\xi=1)=0.4410;$$

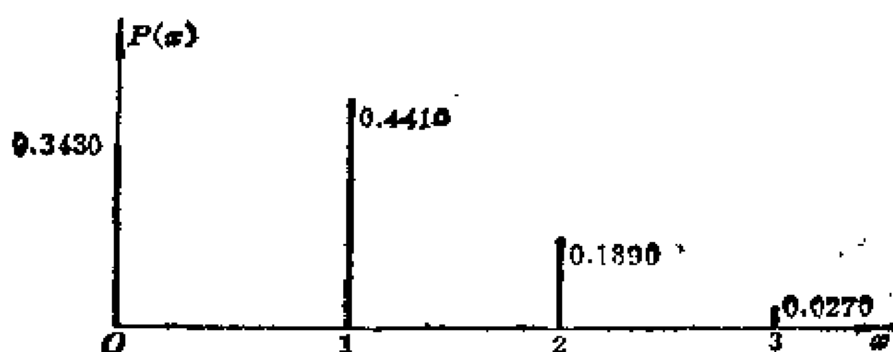


图 2.1.4 $b(3,0.3)$的概率函数 $p(x_k)$图

$$P(\xi=2)=0.1890;$$

$$P(\xi=3)=0.0270.$$

二项分布可以作为描绘射手射击 n 次，其中有 k 次"中的"($k=0,1,\cdots,n$)的概率分布情况的一个数学模型. 或作为随机抛掷硬币 n 次落地时出现 k 次"正面"的概率分布情况的数学模型. 当然也可作为从一批足够多的产品中任意抽取 n 件，其中有 k 件"废品"的模型. 因此，由 §1.5 知它可由贝努里概型产生. 顺便指出，两点分布就是二项分布在 $n=1$ 的特殊情形，单点分布就是二项分布在 $n=0, p=0$ 的特殊情形，(其中设 $c=0$).

表 2.1.1 分别给出了对于 $p=0.2$ 及 $n=9, 16, 25$ 的二项分布 $b(9,0.2), b(16,0.2), b(25, 0.2)$的值. 图 2.1.5 是其相应的图形. 在表 2.1.1 的第一列内列出了数值 $x=0, 1, \cdots, 25$; 第二列给出了随机变数取各个 x 值的概率 $f(x)=P(\xi=x)$; 第三列给出了随机变数取所有小于或等于 x 值的累加概率 $F(x)=\sum_{x_k\leqslant x}f(x_k)$.

细心的读者可能会发现，表 2.1.1 中有不一致的地方. 例如，由第一列与第二列可得：$F(6)=0.9998$, $f(7)=0.0003$. 但是，$F(7)\neq 0.9998+0.0003$. 似乎与 $F(x)$的定义要求 $F(7)=F(6)+f(7)$矛盾. 类似这样的情况，在第三与第四列中，或在第五与第六列中，也是存在的. 究其原因，主要是由于表 2.1.1 并不是绝对精

表 2.1.1　二项分布数值表 ($p=0.2$)

x	$n=9$		$n=16$		$n=25$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
0	.1342	.1342	.0281	.0281	.0038	.0038
1	.3020	.4362	.1126	.1407	.0236	.0274
2	.3020	.7382	.2111	.3518	.0708	.0982
3	.1762	.9144	.2463	.5981	.1358	.2340
4	.0661	.9805	.2001	.7982	.1867	.4207
5	.0165	.9970	.1201	.9183	.1960	.6167
6	.0028	.9998	.0550	.9733	.1633	.7800
7	.0003	1.0000	.0197	.9930	.1108	.8908
8	.0000	……	.0055	9986	.0623	.9531
9	.0000	……	.0012	.9997	.0294	.9825
10	……	……	.0002	.9999	.0118	.9948
11	……	……	.0000	.9999	.0040	.9983
12	……	……	.0000	.9999	.0012	.9995
13	……	……	……	.9999	.0003	.9998
14	……	……	……	.9999	.0001	.9999
15	……	……	……	.9999	.0000	.9999
16	……	……	……	1.0000	.0000	.9999
…	……	……	……	……	……	.9999
…	……	……	……	……	……	.9999
…	……	……	……	……	……	.9999
25	.0000	1.0000	.0000	1.0000	.0000	1.0000

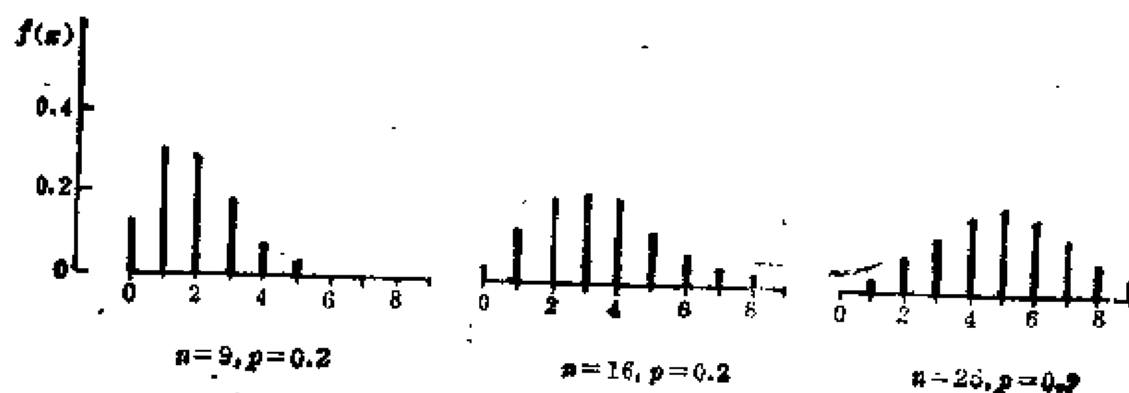


图 2.1.5　二项分布 $b(n,p)$ 图

确的，而是由将真值的小数后面第五位数字加以四舍五入后而得到，更详细地说，考察表 2.1.1 可知，其单数列 $f(x)$ 之值是分别由将其真值四舍五入得出. 双数列 $F(x)$ 之值则是由前一列 $f(x)$ 之值逐项相加后得出$\left(F(x)=\sum_{x_i\leqslant x}f(x_i)\right)$；直至最后不能直接由前一列 $f(x)$ 之值去求（因为积累误差已增大至出现矛盾），再参考 $f(x)$ 原来未经四舍五入的数值，去直接得出 $F(x)$ 的更为精确的近似值.

例 2.1.7 （能量供应问题）设有 9 个工人间歇性地使用电力，在任一时刻每一个工人以同样的概率 0.2 需要一个单位的电力（譬如说，1 个小时中平均有 12 分钟需要电力），若各工人相互独立地工作，问在同一时刻有七个或七个以上的工人需要得到一个单位电力供应的概率是多少？

解 显然这是一个贝努里概型问题，服从二项分布 $b(9,0.2)$. 设 ξ 为在同一时刻需要供应一个单位电力的工人数，有

$$P(\xi\geqslant 7)=\sum_{j=7}^{9}C_9^j(0.2)^j(0.8)^{9-j}=0.0003.$$

上述数据可由查表 2.1.1. 的第一列得到. 上述结果说明，若最多只能供应 6 个单位电力，则超负荷的概率为 0.0003. 或者说，平均地在 3333 分钟中，只有 1 分钟超负荷. 亦即约 55.5 个工作时中可能有一分钟超负荷.（注意：此题的解法也可由查表 2.1.1 之第二列来求得. 事实上，

$$\begin{aligned}P(\xi\geqslant 7)&=1-P(\xi<7)\\&=1-F(6)=1-0.9998\\&=0.0002.\end{aligned}$$

这略为不同于上述答案. 我们在这里建议去查表 2.1.1 之第一列. 原因是：在现在的情形，查第一列时，只用到三个数据：

$P(\xi=7)+P(\xi=8)+P(\xi=9)$，其积累误差会相对地比查第二列时小.)

例 2.1.8 （药效试验）设某种鸭在正常情况下感染某种传染病的概率为20%，现新发明两种疫苗，疫苗A注射在9只健康鸭后无一只感染传染病，疫苗B注射在25只鸭后仅有一只感染，试问应如何评价这两种疫苗，能否初步估计那种较为有效？

解 若疫苗A完全无效，则注射后鸭受感染的概率仍为0.2，故9只鸭中无一只感染的概率为$(0.8)^9=0.1342$(见表2.1.1第一列). 同理，若疫苗B完全无效，则25只鸭中至多有一只感染的概率为

$$(0.8)^{25}+C_{25}^1(0.2)^1\cdot(0.8)^{24}=0.0274.$$

(见表2.1.1最后一列). 因为概率0.0274很小，并且比概率0.1342小得多，因此，可以初步认为疫苗B是有效的，并且比疫苗A有效.

从图2.1.5可以看出，二项分布具有如下的性质:

1) 对于固定的n和p，ξ取k的概率随着k的增大起先增大，直至达到最大值，然后再下降；

2) 对于固定的p，随着n的增大，$b(n,p)$的图形趋于对称.（设$0<p<1$）

若记二项分布$b(n,p)$的通项为

$$b(k;n,p)=C_n^k p^k(1-p)^{n-k}\quad(k=0,1,\cdots,n),$$

则有

$$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{(n-k+1)p}{kq}=1+\frac{(n+1)p-k}{kq},$$

故当$k<(n+1)p$时，$b(k;n,p)$大于前一项，亦即随着k的增加而上升；当$k>(n+1)p$时，则下降；当$(n+1)p=m$为正整数时，两项相等，即$b(m;n,p)=b(m-1;n,p)$，此时该两项为最大值. 若$(n+1)p$不是正整数，则满足$(n+1)p-1<m\leqslant(n+1)p$之正整

数 m，使 $b(m;n,p)$ 为最大值. 这说明：二项分布的通项 $b(k;n,p)$ 当 k 由 0 变到 m 时，单调上升；当 $k=m$ 时达到最大值（若 $(n+1)p$ 为正整数，则有两个最大值 $b((n+1)p;n,p)$ 与 $b((n+1)p-1;n,p)$. 否则，只有一个最大值 $b(m;n,p)$，其中 m 满足：$(n+1)p-1<m\leqslant(n+1)p$）.

当 k 继续由 m 变到 n 时，$b(k;n,p)$ 下降. 常常称 m 为最可能出现次数，称 $b(m;n,p)$ 为中心项. 现举几例说明它的应用.

例 2.1.9 在例 2.1.7 有关能量供应问题中，求最大可能有多少个工人同时需要得到一个单位的电力供应？

解 由于 $n=9,p=0.2$，故

$$(n+1)p=2$$

为正整数，从而知最大可能是 1 个工人或 2 个工人同时要求供应一个单位电力. 这从表 2.1.1 第一列也可直接看出来. 在这种情况下的概率为 $b(2;9,0.2)=0.3020$.

例 2.1.10 在例 2.1.8 有关药效试验问题中，求在正常情况下，没有注射疫苗时 9 只健康鸭与 25 只健康鸭当中分别最可能受到感染传染病的鸭数.

解 对于 9 只健康鸭的情形. 因为 $n=9$，$p=0.2$，由上例可知，最可能是有 1 只或 2 只鸭受感染.

对于 25 只健康鸭的情形，因为 $n=25,p=0.2$，故

$$(n+1)p=5.2$$

不是正整数，故可知在这种情况下，最可能有 5 只鸭受感染. 在这种情形下的概率分别为 $b(2;9,0.2)=0.3020$，及 $b(5;25,0.2)=0.1960$. 也可由表 2.1.1 看出.

从图 2.1.5 说明的 $b(n,p)$ 随着 n 增大而趋于对称这一现象，使我们想到 n 的改变，可能引起二项分布趋于某种极限分布. 在不同的条件下，可以设想会得到不同的极限分布. 下面的泊松定

理就是说明二项分布的泊松逼近. 设有一串正数 p_n, $0<p_n<1$ $(n=1,2,\cdots)$, 我们有

定理 2.1.1 泊松定理①) 若 $\lim\limits_{n\to\infty} np_n=\lambda\geqslant 0$, 则

$$\lim_{n\to\infty} b(k;\ n,p_n)=\lim_{n\to\infty} C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$$

$$(k=0,1,\cdots,n,\cdots)②. \qquad (2.1.2)$$

证 当 $k\geqslant 1$ 时,

$$b(k;n,p_n)=\frac{n(n-1)\cdots(n-k+1)}{k!}p_n^k(1-p_n)^{n-k}$$

$$=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n\cdot\frac{n-k}{n}},$$

其中

$$\lambda_n=np_n.$$

显然

$$\lim_{n\to\infty}\lambda_n^k=\lambda^k;\ \lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^n=e^{-\lambda};$$

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1, \text{(对任意 } k \text{ 成立)}$$

故

$$\lim_{n\to\infty} b(k,n,p_n)=\frac{\lambda^k}{k!}e^{-\lambda}.$$

当 $k=0$ 时, 显然 $b(0;n,p_n)=\left(1-\frac{\lambda_n}{n}\right)^n\to e^{-\lambda}$. ∎

推论 若 $np_n=\lambda\geqslant 0$ $(n=1,2,\cdots)$, 则(2.1.2)成立.

定理 2.1.1 及其推论说明: 若 np_n 恒等于常数 λ, 或 p_n 足够

---

① S. D. Poisson (1781—1840), Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités, 1837.

② 规定 $C_n^0=1, 0!=1$.

小，n 足够大(希望使得 $\lim\limits_{n\to\infty} np_n$ 存在)，则一般说来 $b(k;\ n, p_n)\approx e^{-\lambda_n}\frac{\lambda_n^k}{k!}$，其中 $\lambda_n=np_n\quad(k=0,1,\cdots,n)$.

图 2.1.6 及表 2.1.2 说明了这种近似的情况. 这里把(2.1.2)式的右边称为泊松分布，并把整个数列 $\left\{e^{-\lambda}\frac{\lambda^k}{k!};\ k=0,1,\cdots,\right\}$ 记为 $P(\lambda)$，又记为：

$$P(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}\quad(k=0,1,\cdots). \tag{2.1.3}$$

表 2.1.2

k	$b(k;500,0.002)$	$P(k;1)$
0	.3675	.3679
1	.3682	.3679
2	.1841	.1839
3	.0612	.0613
4	.0153	.0153
5	.00303	.00307
6	.00050	.00051
7	.000071	.000073

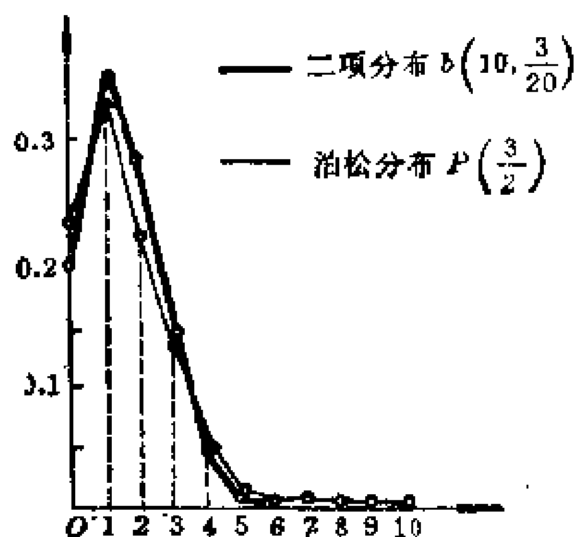


图 2.1.6 二项分布的泊松近似

二项分布的泊松近似，常常被应用于研究稀有事件(即每次试验中事件出现的概率 p 很小)，当贝努里试验的次数 n 很大时，事件发生的频数的分布. 实际表明，在一般情况下，当 $p<0.1$ 时，这种近似是很好的，甚至不必 n 很大都可以(这点从比较二项分布与泊松分布的概率分布表也可以看出). 例如，当 $p=0.01$ 时，甚至 $n=2$ 时，这种近似程度就已经很好了. 下表说明了这一情况，其中 $np=0.02$.

k	$b(k;2,0.01)$	$P(k;0.02)$
0	0.9801	0.9802
1	0.0198	0.0196
2	0.0001	0.0002

现举例说明泊松定理的应用，并考虑产生的误差.

例 2.1.11 设枪击飞机，每次命中目标的概率为 $p=0.001$，若射击 5000 次，试求击中两弹或两弹以上的概率.

解 令 ξ 为击中的弹数，显然 ξ 服从二项分布，故

$$b(0;5000,0.001)=(0.999)^{5000}=0.0071.$$

由于此时 $\lambda=np=5$，故由泊松分布算得

$$P(0;5)=e^{-5}\approx 0.006738.$$

又

$$b(1;5000,0.001)=(0.999)^{4999}\cdot(0.001)\cdot 5000=0.0354,$$

而

$$P(1;5)=e^{-5}\cdot 5\approx 0.03369.$$

故所求的概率等于

$$P(\xi\geqslant 2)=1-b(0;5000,0.001)-b(1;5000,0.001)$$
$$=0.9575,$$

而用泊松分布时，

$$P\{\xi\geqslant 2\}=1-P\{0;5\}-P\{1;5\}\approx 0.959572.$$

可见用泊松定理计算所得的误差在此例中小于 0.25%. 易知，最可能击中次数为 5，其概率 ≈ 0.175468.

上面二项分布概率的计算，可由取对数后直接计算①. $P(k;\lambda)$ 则由表查得②(例如当 $\lambda\leqslant 4.4$ 时，本书的附录中有表可查).

例 2.1.12 设一女工照管 800 个纱锭，若每一纱锭单位时间内纱线被扯断的概率为 0.005，试求最可能的扯断次数及概率；又求单位时间内扯断次数不大于 10 的概率.

---

① 当 n 小于 50 时，二项分布有表可查. 例如：Tables of the Binomial Probability Distribution, the National Bureau of Standards, New York, 1950.

② E. C. Molina, Poisson's Exponential Binomial Limit, D. Van Nostrand Company, Inc, Princeton, N. J., 1942.

解 现在 $n=800, p=0.005, np=4, (n+1)p=4.005$. 故最可能扯断次数为 4. 其概率为 $b(4;800,0.005)=C_{800}^{4}\cdot 0.005^{4}\cdot 0.995^{796}=0.1945$. 用泊松分布时算得 $P\{4;4\}=0.195367$.

在单位时间内扯断次数不大于 10 的概率为

$$\sum_{k=0}^{10} b(k;\ 800,0.005)=1-\sum_{k=11}^{800} b(k;800,0.005)$$

$$\approx 1-\sum_{k=11}^{\infty} P(k;4)=1-0.002840=0.997160.$$

此例说明，利用泊松定理近似计算二项分布的概率值是很方便的. 否则要用对数直接计算二项分布的概率，将带来很大的计算量.

例 2.1.13 设根据过去的统计，已知在某种产品中出现废品的概率为 $p=0.014$，现若要求有 90% 的可能性在一个箱子的这种产品中能选得 100 个合格产品，试问在一个箱子中至少应放多少个产品？

解 设 $100+x$ 为所求的产品数，则问题变为去求最小的 x 使 $b(0;\ 100+x,0.014)+b(1;\ 100+x,0.014)+\cdots+b(x;\ 100+x,0.014)\geqslant 0.9$. 由于 $np\approx 100\times 0.014=1.4$，利用泊松近似可去求最小的 x，使

$$P(0;1.4)+P(1;1.4)+\cdots+P(x;1.4)\geqslant 0.9.$$

查表可得，当 $x=3$ 时，左边 ≈ 0.9463. 故可知只需在箱子中放进 103 个产品，就可保证有 90% 的可能性使其中有 100 个是合格品.

下面讨论泊松分布的性态.

例 2.1.14 若随机变数 ξ 由下表给出(其中 $\lambda>0$ 为常数)

ξ	0	1	2	$\cdots$	k	$\cdots$
$P(\xi=k)$	$e^{-\lambda}$	$e^{-\lambda}\frac{\lambda}{1!}$	$e^{-\lambda}\frac{\lambda^2}{2!}$	$\cdots$	$e^{-\lambda}\frac{\lambda^k}{k!}$	$\cdots$

由前知 ξ 服从(参数为 λ 的)泊松分布 $P(\lambda)$.

可以证明上表确满足(2.1.1)式. 首先, $P(\xi=k)\geqslant 0$ 是显然的. 其次,由

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1$$

可知 $\sum_{k=0}^{\infty} P(\xi=k)=1$.

图 2.1.7 与 2.1.8 分别给出了泊松分布的(概率)质量分布图与概率函数图(在 $\lambda=3.5$ 的情形).



图 2.1.7　泊松分布的质量图($\lambda=3.5$)

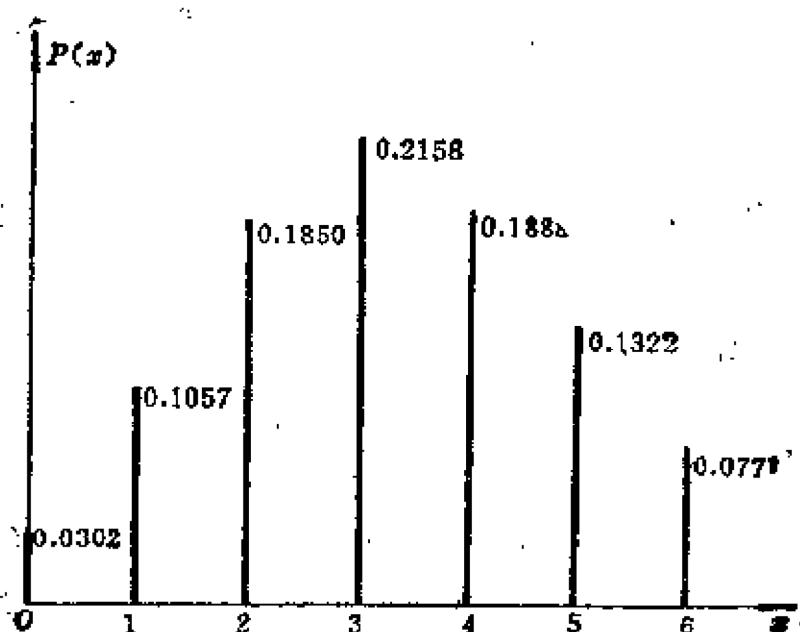


图 2.1.8　泊松分布的概率函数图($\lambda=3.5$)

$$P(\xi=0)=0.0302;$$

$$P(\xi=1)=0.1057;$$

$$P(\xi=2)=0.1850;$$

$$P(\xi=3)=0.2158;$$

$$P(\xi=4)=0.1888;$$

$$P(\xi=5)=0.1322;$$

$$P(\xi=6)=0.0771;$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$P(\xi=13)=0.0001.$$

图 2.1.9 表示了不同参数 λ 所对应的泊松分布图. 在$\lambda\leqslant 1$时，图形明显地向一边偏斜；当 λ 增加时，图形变得接近对称.

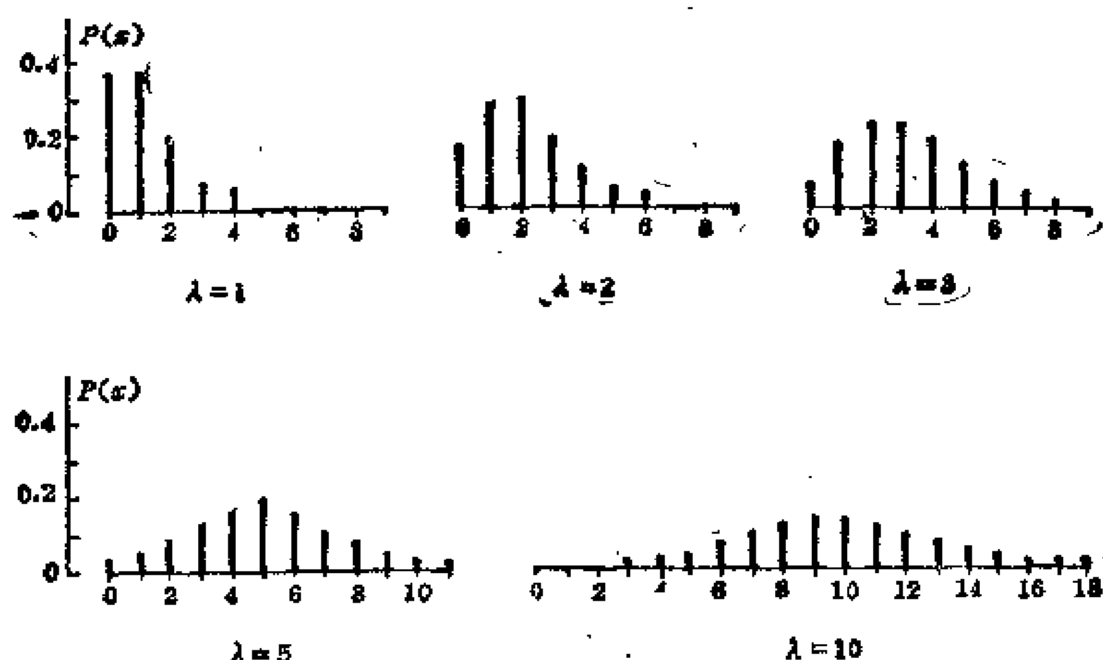


图 2.1.9　不同参数 λ 所对应的泊松分布图

由泊松定理可知，泊松分布可以作为描绘大量试验中稀有事件出现的频数 $k=0,1,\cdots$的概率分布情况的一个数学模型. 诸如：飞机被击中的子弹数；纱锭的纱线被扯断的次数；大量螺钉中不合格品出现的次数；一个集团中生日是元旦的人数；三胞胎出生的次数；一年中死亡的百岁老人数；一页中印刷错误出现的数目；一年中暴雨出现在夏季中的次数；数字通讯中传输数字时发生误码的个数；…等等随机变数，就都相当近似地服从泊松分布.

例 2.1.15　自 1875 年到 1955 年中的某 63 年间，上海夏季

(5—9 月间）共发生暴雨 180 次，试求在一个夏季中发生 k 次($k=0,1,2,\cdots 8$)暴雨的概率 P_k(设每次暴雨以 1 天计算)。

解 一年夏天共有的天数为

$$n=31+30+31+31+30=153,$$

故可知发生暴雨的概率约为 $p=\frac{180}{63\times 153}$，这值甚小，而 $n=153$ 则较大. 现用泊松近似. 由于

$$\lambda=np=\frac{180}{63}\approx 2.9.$$

因此，可得

$$P_0=P(0;2.9)=e^{-\lambda}=e^{-2.9}\approx 0.055;$$

$$P_1=P(1;2.9)=\lambda e^{-\lambda}=\lambda P_0=0.160;$$

$$P_2=\frac{\lambda}{2}P_1=0.2320;$$

$$P_3=\frac{\lambda}{3}P_2=0.2243;$$

$$P_4=\frac{\lambda}{4}P_3=0.1626;$$

$$P_5=\frac{\lambda}{5}P_4=0.0943;$$

$$P_6=\frac{\lambda}{6}P_5=0.0456;$$

$$P_7=\frac{\lambda}{7}P_6=0.0189;$$

$$P_8=\frac{\lambda}{8}P_7=0.0068;$$

一般地，

$$P(k;\lambda)=\frac{\lambda}{k}\left(\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}\right)=\frac{\lambda}{k}P(k-1;\lambda). \qquad (2.1.4)$$

一个夏季中发生 k 次暴雨的概率为 p_k，故 63 年中应有 $63p_k$

个夏季发生k次暴雨．下表是实际观察值与理论计算值的对照表，例如，发生二次暴雨的，实际上有14个夏季，计算则为14.8；总起来看，符合情况较好．

①

暴雨次数	0	1	2	3	4	5	6	7	8
实际年数	4	8	14	19	10	4	2	1	1
理论年数	3.5	10.1	14.6	14.1	10.2	6	2.9	1.2	0.42

①么枕生，气候统计，科学出版社，1963.

泊松分布固然可以作为大量试验中稀有事件出现的频数的概率分布的数学模型，但是理论与实践都说明，一般说来它也可作为下列随机变数的概率分布的数学模型：

在任给一段固定的时间间隔内，

1）由某块放射性物质放射出的α-质点，到达某个计数器的质点数；

2）从一个真空管的阴极发射出的电子，到达阳极的电子数；

3）来到某公共设施要求给予服务的顾客数；这里的公共设施的意义，可以是极为广泛的．诸如百货商店的售货员；工厂的仓库保管员；机场的跑道（在这种情形，顾客可以相应地想像为飞机．其它情形亦类似）；港口的装卸货物设备；金工车间的修理工；电话交换台的干线等等（在这里，让我们假设在每一个极短时间内，同时到来要求服务的顾客至多只有一个．在很多情况下，会是这样的．但是，实践证明，也有同时一批顾客一起到来要求服务的情况发生．这也是值得重视与研究的课题．事实上，排队论已发展出新的方法来进行处理与研究这类问题．但由于它已超越出本书的范围，我们在此不去讨论它）；

4）事故、错误、故障及其它灾害性事件数．

这些随机变数大致上都有如下的特点：它们都取正整数为值，

并且与时间间隔长度有关①. 当时间间隔极短，取值为数 2 以上是几乎不可能的(例如，在极短的时间间隔内，可以认为不能有两个或两个以上的电话呼叫同时来到). 另外，它们取值的概率，只与时间间隔的长度有关，而与从哪个时刻开始算起没有什么关系. 并且，在不相重叠的时间间隔内，彼此没有什么影响. 我们可以证明，在满足上述相应的数学条件下，这种与时间有关的随机变数，确实是服从泊松分布的. 这正是大量自然科学中的随机现象可以用泊松分布进行研究与处理的另一重要根据.

在证明这一论断之前，让我们先引入下述数学分析中的一个结果.

引理 设 $f(t)$ 是定义在 $t>0$ 上的一个实值函数，对任意的 $t_1, t_2>0$ 满足

$$f(t_1+t_2)=f(t_1)\cdot f(t_2), \tag{2.1.5}$$

且 $f(t)$ 在任意的有限区间上有界，若 $f(1)\neq 0$，则必存在常数 r，使

$$f(t)=e^{-rt} \quad (t>0). \tag{2.1.6}$$

*证 由 $f(t_1+t_2)=f(t_1)\cdot f(t_2)$，易得对任意正整数 n，有 $f(nt)=[f(t)]^n$. 取 $t=\frac{1}{n}$，则有 $f(1)=\left[f\left(\frac{1}{n}\right)\right]^n$. 因此，对任意正整数 m, n，有

$$f\left(\frac{m}{n}\right)=[f(1)]^{\frac{m}{n}}. \tag{2.1.7}$$

令 $r=-\ln f(1)$(因为 $f(1)=\left[f\left(\frac{1}{2}\right)\right]^2>0$)，

可得

$$f\left(\frac{m}{n}\right)=e^{-r\cdot\frac{m}{n}}. \tag{2.1.8}$$

---

① 其实，相应于不同的时间间隔，我们得到的是不同的随机变数.

剩下只需证：对任意的无理数 $t>0$，有

$$f(t)=e^{-\gamma t}$$

我们用反证法证明上述结论.

若存在无理数 t_0，使 $f(t_0)\neq e^{-\gamma t}$（亦即 $f(t_0)\neq[f(1)]^{t_0}$）.不妨设 $t_0>1$（事实上，若对一切 $t>1$，得证

$$f(t)=e^{-\gamma t};$$

则当 $0<t\leqslant 1$ 时，可有

$$f(t)\cdot f(1)=f(t+1)=e^{-\gamma(t+1)}=e^{-\gamma t}\cdot e^{-\gamma}$$

亦即 $f(t)=e^{-\gamma t}$ 仍真）. 现记

$$F(t)=f(t_0t)/[f(t_0)]^t \tag{2.1.9}$$

（因为 $f(t_0)=\left[f\left(\frac{t_0}{2n}\right)\right]^{2n}\geqslant 0$，其中 $t_0/2n<1$，n 为自然数；并且 $f(1)=f\left(\frac{t_0}{2n}\right)\cdot f\left(1-\frac{t_0}{2n}\right)>0$. 故(2.1.9)有意义），易知 $F(t)$ 仍满足(2.1.5)，故也满足(2.1.7). 因此，若证明了 $F(t_0^{-1})=1$，则证明了下面两等式：

$$1=F(t_0^{-1})=f(1)/[f(t_0)]^{t_0^{-1}},$$

$$f(t_0)=[f(1)]^{t_0}=e^{-\gamma t_0}.$$

这就得出了矛盾，因而(2.1.6)式成立.

下面证明 $F(t_0^{-1})=1$.

仍用反证法. 设 $F(t_0^{-1})\neq 1$，则或者有 $F(t_0^{-1})>1$，或者有 $F(1-t_0^{-1})>1$（因为 $F(t_0^{-1})F(1-t_0^{-1})=F(1)=1$）.

令 τ 为 t_0^{-1} 或 $1-t_0^{-1}$，则有

$$F(N\tau)=[F(\tau)]^N\longrightarrow\infty\quad(N\to\infty).$$

然而，可以证明，对任意 $N=1,2,\cdots$

$$|F(N\tau)|\leqslant M\quad(M\text{ 为某一实常数})$$

这与 $F(N\tau)=[F(\tau)]^N\to\infty$ 矛盾. 因而得 $F(t_0^{-1})=1$.

事实上，对任意 $t\in R_1$，

$$|F(t)|\leqslant M,$$

这是因为对任一 $t\in R_1$，必存在 $t'(0<t'\leqslant 1)$ 及有理数 r，使得 $t=r+t'$，而由假设有

$$|F(t)|=|F(r)F(t')|=|F(t')|$$
$$=|f(t_0t')/[f(t_0)]^{t'}|\leqslant M. \blacksquare$$

推论 1 设 $f(t)$ 满足(2.1.5)及在任意有限区间上有界，若存在 t_0，使 $f(t_0)\neq 0$，则必存在常数 r，能够使(2.1.6)成立.

证 考虑 $F(t)$ 如(2.1.9)式，则对于 $F(t)$ 可应用引理，故存在 r_0'，使

$$\frac{f(t_0t)}{(f(t_0))^t}=F(t)=e^{-r_0't},\quad (t>0).$$

亦即：对一切 $t>0$，有

$$f(t_0t)=(f(t_0))^te^{-r_0't}=e^{-\frac{(r_0'-\ln f(t_0))}{t_0}\cdot t_0t} \blacksquare$$

推论 2 若 $f(t)$ 满足(2.1.5)式并且连续或单调，又不恒为 0，则(2.1.6)成立.

下面我们来论证上面提到的有关产生泊松分布的论断. 为了易于理解，我们采用有关电话呼叫的术语来叙述.

以 ξ_t 表示在时间区间 $(0,t]$ 上来到的电话呼叫数，若它满足下述条件，则称它为一个泊松过程或泊松流：

1°平稳性. 在 $(t_0,t_0+t]$ 上来到的呼叫数是 k 的概率 $P_k(t)=P(\xi(t_0,t_0+t]=k)$① 只与时间区间的长度 t 有关而与时间起点 t_0 无关 $(k=0,1,\cdots)$；

2° 独立增量性（无后效性）. 在任意 n 个不相交的时间区间 $(a_i,b_i](i=1,2,\cdots,n)$ 中，各自来到的呼叫数是相互独立的；即若第 i 个时间区间 $(a_i,b_i](i=1,2,\cdots,n)$ 有 k_i 个呼叫来到，则事件 $\xi(a_i,b_i]=k_i(i=1,2,\cdots,n,)$ 是独立的；

---

① 以 $\xi(t_0,t_0+t]$ 表示在时间区间 $(t_0,t_0+t]$ 上来到的呼叫次数.

3° 普通性. 在足够小的时间区间中,最多来到一个呼叫; 即若记 $\psi(t)=1-P_0(t)-P_1(t)$,则

$$\lim_{t\to 0}\frac{\psi(t)}{t}=0,$$

故可记

$$\psi(t)=o(t);$$

4° 非平凡情形. $P_0(t)\neq 1$, $\sum_{k=0}^{\infty}P_k(t)=1$. ①

性质 1° 说明泊松过程的概率分布不会因时间的推移而发生改变. 性质 2° 说明前一时间区间有无呼叫或来到多少个呼叫并不影响到后面时间区间上有多少呼叫发生的概率. 性质 3° 说明相对于时间长度 t,当 t 很小时,二次以上的呼叫来到的可能性是可以忽略的. 性质 4° 说明不会在任何时间区间中恒无呼叫来到,并且在任何时间区间中必以正概率发生如下事件:或者没有呼叫、或者来到 1 个呼叫、… 等等 (最后一点, 其实是由于 ξ_t 为随机变数并且只取 0,1,2,…为值所要求的).

对于泊松过程,成立下面的定理:

定理 2.1.2 泊松过程 ξ_t 必服从以 λt 为参数的泊松分布 $P(\lambda t)$,其中 $\lambda>0$ 是常数,即

$$P(\xi_t=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\quad(k=0,1,2,\cdots). \tag{2.1.10}$$

证 由 1° 与 2°,易得 $P_0(t)$满足

$$P_0(t_1+t_2)=P_0(t_1)\cdot P_0(t_2).$$

显然 $P_0(t)$是单调下降的,并且 $P_0(t)$ 不能恒等于 0(否则,在任意的时间区间中必有呼叫来到. 从而可得在任一有限区间中必有无

① 按照定义与概率意义理解,$P_0(0)=P(\xi(0,0]=0)$表示在时间长度为零的集合上来到零次呼唤;这当然可看作一个必然事件,故自然地有 $P_0(0)=1$.

有多个呼叫来到，这便与假设 4° 矛盾），故由引理推论 2 可得，存在正常数 $\lambda>0$①，使

$$P_0(t)=e^{-\lambda t}\quad(t>0).\tag{2.1.11}$$

往下求 $P_k(t)$. 将 $(0,t]$ 分成 n 个等长的互不相交的子区间，且 $n>k$. 记 A_1 为“在某 k 个子区间各刚好出现一个呼叫，而在其余的 $n-k$ 个子区间中没有呼叫来到”这一随机事件；A_2 为“至少有一个子区间出现一个以上的呼叫”；A_3 表“在 $(0,t]$ 中有 k 个呼叫来到”. 则

$$P_k(t)=P(A_1)+P(A_2\cap A_3),\tag{2.1.12}$$

$$P(A_1)=C_n^k\left[P_1\left(\frac{t}{n}\right)\right]^k\left[P_0\left(\frac{t}{n}\right)\right]^{n-k},$$

且

$$\left[P_0\left(\frac{t}{n}\right)\right]^{n-k}=e^{-\lambda t\cdot\frac{n-k}{n}}=e^{-\lambda t}\cdot e^{\frac{k\lambda t}{n}}=e^{-\lambda t}[1+o(1)].$$

$$\left[P_1\left(\frac{t}{n}\right)\right]^k=\left[1-e^{-\frac{\lambda t}{n}}-\psi\left(\frac{t}{n}\right)\right]^k=\left[1-e^{-\frac{\lambda t}{n}}+o\left(\frac{1}{n}\right)\right]^k$$

$$=\frac{(\lambda t)^k}{n^k}[1+o(1)].$$

故

$$P(A_1)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}[1+o(1)].$$

从而

$$P(A_1)\longrightarrow e^{-\lambda t}\cdot\frac{(\lambda t)^k}{k!}\quad(n\to\infty).\tag{2.1.13}$$

注意到

$$P(A_2\cap A_3)\leqslant P(A_2)\leqslant n\psi\left(\frac{t}{n}\right)=\frac{\psi\left(\frac{t}{n}\right)}{\frac{t}{n}}\cdot t\to 0\quad(n\to\infty),\tag{2.1.14}$$

① 因为当 $t>0$ 时，$0\leqslant P_0(t)\leqslant 1$，故可知 $\lambda\geqslant 0$，再由 $P_0(t)\not\equiv 1$，故得 $\lambda>0$.

由(2.1.13)及(2.1.14)，在(2.1.12)中令 $n\to\infty$，立即可得

$$P_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}. \quad \blacksquare$$

泊松分布既然在适当条件下，可由二项分布取极限得到。因此，自然会问：二项分布所具有的某些性质，泊松分布能否同样具有？例如，关于分布图形的上升下降性及最大值等。

由

$$\frac{P(k;\lambda)}{P(k-1;\lambda)}=\frac{\lambda}{k} \tag{2.1.15}$$

可见，当 $k<\lambda$ 时，$P(k-1;\lambda)<P(k;\lambda)$ 当 $k>\lambda$ 时，$P(k-1;\lambda)>P(k;\lambda)$。在 λ 是正整数时，有 $P(\lambda;\lambda)=P(\lambda-1;\lambda)$。故得结论如下：泊松分布的通项 $P(k;\lambda)$ 当 k 由 0 变到 $[\lambda]$（这里以 $[\lambda]$ 记不超过 λ 的最大整数）时，单调上升，并且在 $k=[\lambda]$ 时，达到最大值 $P([\lambda];\lambda)$（若 $\lambda=[\lambda]$，则有两个最大值 $P(\lambda;\lambda)=P(\lambda-1;\lambda)$）；当 k 超过 λ 继续变动时，$P(k;\lambda)$ 单调下降，常常称 $[\lambda]$ 为最可能出现次数。

易知，如果 $\lim\limits_{n\to\infty}b(k;n,p_n)=P(k;\lambda)$（假设 $\lim\limits_{n\to\infty}np_n=\lambda$，以下恒设 $0<p_n<1$），那么当 n 足够大时，$b(k;n,p_n)\sim P(k;np_n)$ $(k=0,1,2,\cdots,n)$。在这种情况下，由不等式 $(n+1)p_n-1<np_n\leqslant(n+1)p_n$ 及 $\lim\limits_{n\to\infty}np_n=\lambda$，可以推知：当 np_n 与 $(n+1)p_n$ 都不是正整数时，二项分布最可能出现次数为 m，它满足 $(n+1)p_n-1<m<(n+1)p_n$。但是，（以 np_n 为参数的）泊松分布的最可能出现次数却是：$[np_n]=m-1$（当 $[np_n]<(n+1)p_n-1$ 时），或 $[np_n]=m$（当 $[np_n]>(n+1)p_n-1$ 时）或 $[np_n]=m$（当 $[np_n]>(n+1)p_n-1$ 时）。在此情形，二者都只能有一个最可能出现次数，这个泊松分布的最可能出现次数与二项分布的最可能出现次数或者相等或者泊松分布的比二项分布的小 1。但在 np_n 为正整数时，泊松近似的最可能出现次数为 $[np_n]$ 与 $[np_n]-1$；但二项分布 $b(n,p_n)$ 的最可能出现次数却只有一个，它为 $[np_n]$（因为在这种情形下 $(n+1)p_n$ 不是正整数）。同理，在 $(n+1)p_n$ 是正整数时，二项分布的最可能次数有两个：$(n+1)p_n$ 与 $(n+1)p_n-1$；但它的泊松近似却只有一个最可能出现次数 $[np_n]$（因为在这种情形下 np_n 不是正整数），并且 $[np_n]=(n+1)p_n-1$。

例2.1.16 试求在例 2.1.15 中一个夏季内最可能出现的暴雨次数及其概率。

解 在考虑为二项分布模型或泊松分布模型时，其最可能出现的暴雨次数都是$[np]=2$，其出现的概率为$P_2=0.2320$. 在例2.1.15最后计算所得的表中，我们可以看到，按理论计算，果然是出现2次暴雨的理论年数最多(14.8年). 但实测数据则为3次暴雨出现的年数最多(19年)，其次才是出现2次暴雨的年数(14年)，这与理论计算结果略有差异.

还有一类离散型随机变数，也是较为常见的. 并且它的分布列与二项分布有着密切关系. 在一定条件下，后者是它的极限分布.

称随机变数ξ是服从超几何分布的，若它满足：

ξ	0	1	…	k	…	l
$P(\xi=k)$	$\frac{C_M^0 C_{N-M}^n}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	…	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	…	$\frac{C_M^l C_{N-M}^{n-l}}{C_N^n}$

其中$0\leqslant n\leqslant N$，$0\leqslant M\leqslant N$，$l=\mathrm{Min}(M,n)$，并且它们都是非负整数. 由下面的等式：

$$(1+x)^N=(1+x)^M\cdot(1+x)^{N-M},$$

对比x^n之系数，容易得证

$$\sum_{k=0}^{l}C_M^k C_{N-M}^{n-k}/C_N^n=1.$$

故上表第二列确是分布列. 若设想有N个产品，其中M个废品，任意抽取n个，则其中恰有的废品个数ξ，便服从上述的超几何分布. 容易验证：若$\lim\limits_{N\to\infty}\frac{M}{N}=p$，(亦即，在无限多个产品中，废品率是$p$)，则在$n,k$保持不变的条件下，有$\lim\limits_{N\to\infty}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=C_n^k p^k(1-p)^{n-k}$.
亦即，超几何分布的极限分布是二项分布. 这正与我们的直观想象符合：在废品率为确定数p的足够多的产品中，任意抽取n个，(即做n次独立试验)，其中恰有k个废品的情形(即"废品出现"这一事件发生了k次)，由于产品数足够多，不放回与放回无区别(即

废品率可看作不变). 在这种情形下，正是典型的贝努里试验概型. 故可用二项分布去刻划其概率分布列.

现在让我们来证明:

$$\lim_{N\to\infty}\frac{M}{N}=P \Rightarrow \qquad \lim_{N\to\infty}\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=C_n^k p^k(1-p)^{n-k}.$$

事实上,由于

$$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}=\frac{M!}{k!\,(M-k)!}\cdot\frac{(N-M)!}{(n-k)!\,(N-M-n+k)!}$$

$$\cdot\frac{n!\,(N-n)!}{N!}$$

$$=\frac{n!}{k!\,(n-k)!}\cdot\frac{M\cdot(M-1)\cdots(M-k+1)}{N\cdot N\cdots N}$$

$$\cdot\frac{(N-M)\cdots(N-M-(n-k)+1)}{N\cdots N}$$

$$\cdot\frac{N\cdot N\cdots N}{N(N-1)\cdots(N-n+1)}\equiv C_n^k a_N b_N c_N,$$

其中$\lim\limits_{N\to\infty}a_N=p^k$, $\lim\limits_{N\to\infty}b_N=(1-p)^{n-k}$, $\lim\limits_{N\to\infty}c_N=1$. 故得证.

下面进入讨论非离散型随机变数.

二、连续型随机变数及其密度函数

如例 2.1.2 所述，加工的零件长度与规定的长度的偏差值可以取值于包含 0 点的某一区间，对于这一类可以在某一区间内任意取值的随机变数 ξ，由于它的值不是集中有限个或可列无穷个点上;因此,只有确知取值于任一区间上的概率 $P\{a<\xi<b\}$(其中 $a<b$ 为任意实数),才能掌握它取值的概率分布情况.

对于取值非离散的随机变数，其中有一类很重要且常见的类型,就是所谓连续型的随机变数.

定义 2.1.3 若存在非负可积函数 $f(x)$, $\int_{-\infty}^{\infty}f(x)dx<\infty$, 使随机变数 ξ 取值于任一区间(a,b)的概率为

$$P\{a<\xi<b\}=\int_a^b f(x)dx, \tag{2.1.16}$$

则称 ξ 为具有连续型分布或称 ξ 为连续型随机变数. $f(x)$ 称为 ξ 的分布密度函数，有时简称为分布密度或密度函数.

注 在本节之"三"中，引入了分布函数 $F(x)$ 的概念后，可知由(2.1.16)有 $F(x)=\int_{-\infty}^{x} f(t)dt$ 成立. 但学过实变函数的读者知道，此时 $F(x)$ 实际上是绝对连续的. 因此，把这种随机变数称之为绝对连续型，而把分布函数 $F(x)$ 连续，但是又不能以非负函数的变动上限的积分表出的，称之为连续型可能更为恰当①. 考虑到本书并不讨论后者，并且大多数书中都称满足(2.1.16)的随机变数为连续型. 故虽然不是很恰如其分，我们在本书中仍采用这一传统的称呼与定义.

由定义 2.1.3，有

$$\int_{-\infty}^{\infty} f(x)dx=P\{-\infty<\xi<\infty\}=1.$$

因此，同离散型随机变数的概率函数一样(比较(2.1.1)式)，对密度函数 $f(x)$，有

$$\left.\begin{aligned} &f(x)\geqslant 0, \\ &\int_{-\infty}^{\infty} f(x)dx=1. \end{aligned}\right\} \tag{2.1.17}$$

如果我们把 $P\{a<\xi<b\}$ 看作在区间 (a,b) 上的质量，则 $f(x)$ 表示密度.

由定义 2.1.1. 的后面用小字印的一段叙述，知事件 $\{a<\xi<b\}\in\mathscr{F}$. 故 $P\{a<\xi<b.\}$ 有意义.

对于连续型随机变数 ξ，它取任一值 x 的概率为 0. 事实上，对任一 x

$$\{\xi=x\}=\bigcap_{n=1}^{\infty}\left\{\xi<x+\frac{1}{n}\right\}-\{\xi<x\},$$

① A. N. Shiryayev, Probability, Springer-Verlag, N. Y. (1984).

故

$$P\{\xi=x\}=\lim_{n\to\infty}P\left\{\xi<x+\frac{1}{n}\right\}-P\{\xi<x\}$$

$$=\lim_{n\to\infty}\int_{-\infty}^{x+\frac{1}{n}}f(x)dx-\int_{-\infty}^{x}f(x)dx=0.$$

由此可得，对于连续型随机变数 ξ，有

1) $P\{x_1<\xi<x_2\}=P\{x_1\leqslant\xi<x_2\}=P\{x_1<\xi\leqslant x_2\}=P\{x_1\leqslant\xi\leqslant x_2\}$；

2) 概率为 0 的事件不一定是不可能事件. 同样，概率为 1 的事件不一定是必然事件.

对于离散型随机变数 ξ，

ξ	x_1	x_2	…	x_n	…
$P\{\xi=x_n\}$	p_1	p_2	…	p_n	…

并不存在密度函数. 但是若引入 δ-函数（即广义函数，有关它的概念和性质的讨论参考[17]），则对于离散型随机变数，也一样可以有（广义的）密度函数. 例如

$$f_\xi(x)=\sum_{j=1}^{\infty}p_j\delta(x-x_j)$$

就是一个密度函数，它既满足(2.1.17)，也满足(2.1.16).

下面举出一些常用的连续型随机变数的例子.

例 2.1.17 （均匀分布）若随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & \text{当 } a\leqslant x\leqslant b,\\ 0, & \text{其它},\end{cases}\qquad(2.1.18)$$

则称 ξ 服从在 $[a,b]$ 上的均匀分布（或说，ξ 均匀分布于 $[a,b]$）.

可以证明，$f(x)$ 满足(2.1.17).

$f(x)\geqslant0$ 是显然的. 且

$$\int_{-\infty}^{\infty}f(x)dx=\int_a^b\frac{1}{b-a}dx=1.$$

$f(x)$ 的图形如图 2.1.10.

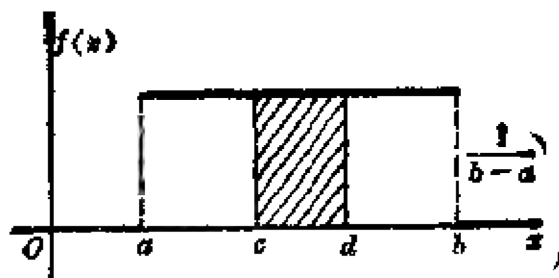


图 2.1.10 均匀分布的密度函数

由(2.1.18)易得

1° $P\{\xi>b\}=\int_b^{\infty}0dx=0$,

同理,$P\{\xi<a\}=0$.

2° 若 $a\leqslant c<d\leqslant b$,则

$$P\{c<\xi<d\}=\int_c^d\frac{dx}{b-a}=\frac{1}{b-a}\cdot(d-c).$$

因此,具有均匀分布的随机变数 ξ 的物理意义是:ξ 取值大于 b 或小于 a 的概率为 0,ξ 概率为 1 地在区间 $[a,b]$ 中取值,并且 ξ 的值落入 $[a,b]$ 中任一区间 $[c,d]$ 中的概率与区间 $[c,d]$ 的长度成正比.

在数值计算中,由于四舍五入,小数点后第一位小数所引起的误差 ξ,一般可以看作是一个服从在 $[-0.5,0.5]$ 上的均匀分布的随机变数. 又如在 $[a,b]$ 中随机掷质点,用 ξ 表示此质点的坐标,则一般地,也可把 ξ 看作是一个服从在 $[a,b]$ 上的均匀分布的随机变数.

例 2.1.18 (正态分布)若随机变数 ξ 的密度函数由式

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\quad(x\in R_1)\tag{2.1.19}$$

给出,则称 ξ 服从标准正态分布,简称 ξ 服从正态 $N(0,1)$.

由(2.1.19)可看出 $\varphi(x)\geqslant0$,且

$$\int_{-\infty}^{\infty}\varphi(x)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx=\frac{1}{\sqrt{2\pi}}\cdot\sqrt{2\pi}=1,①$$

故 $\varphi(x)$满足密度函数的要求.

正态 $N(0,1)$的密度函数图形如图 2.1. 11.

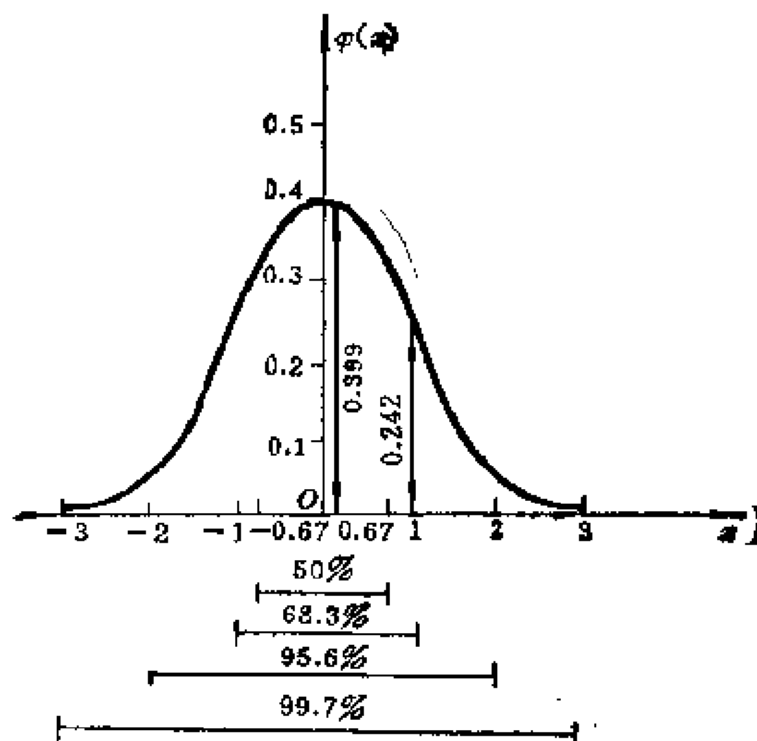


图 2.1.11 正态密度函数 $\varphi(x)$的图. 最下一横线表示: 在$(-3,3)$中曲线之下面积为 99.7%, 其它横线意义类似. 注意全面积为 1.

现考虑函数

$$\varphi_{a,\sigma}(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}\quad(x\in R_1).\tag{2.1.20}$$

显然 $\varphi_{a,\sigma}(x)\geqslant0$, 且

---

① $\left(\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx\right)^2=2\pi$ 的证明如下:

记 $I^2=\left(\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx\right)^2$, 则 $I^2=\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}dx\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dxdy$. 作极坐标变换: $x=r\cos\theta$, $y=r\sin\theta$, 即得 $I^2=\int_0^{2\pi}d\theta\int_0^{\infty}e^{-\frac{r^2}{2}}rdr=2\pi$.

$$\int_{-\infty}^{\infty}\varphi_{a,\sigma}(x)dx=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy=1$$

$\left(\text{其中作了变换}\ y=\frac{x-a}{\sigma}\right)$,故它也是密度函数.

随机变数 ξ,若以(2.1.20)为其密度函数,则称 ξ 服从正态 $N(a,\sigma)$.

正态 $N(a,\sigma)$ 的图形如图 2.1.12,那里的 $a=0$,若 $a\neq 0$,则将图形平移至以 a 点为中心即可.

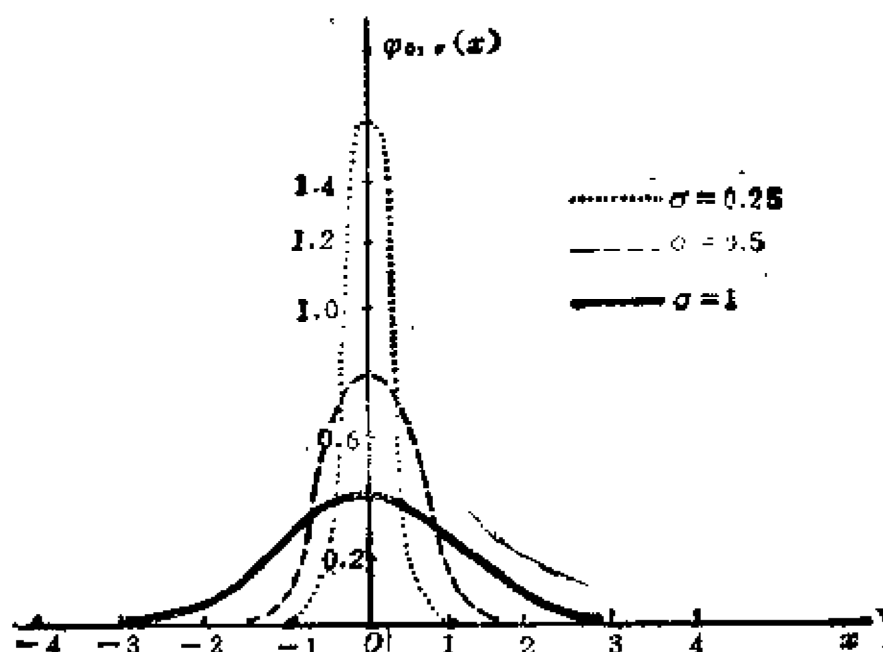


图 2.1.12 正态分布 $N(0,\sigma)$ 图

从图 2.1.11 可见,服从正态 $N(0,1)$ 的随机变数 ξ 的直观意义是:它以很大的概率于 0 的附近取值,而取正值与取负值的可能性(概率)相同,取值于充分大的数的一个邻域内的可能性是很小的.

用式子来表示,就是:若 ξ 服从正态 $N(0,1)$ 则

1° $P(-3<\xi<3)=0.997\approx 1$(此点可查表求得).

2° 对于任意的 $0<a<b$，有

$$P\{a<\xi<b\}=P\{-b<\xi<-a\}$$

3° 当 $0<M$ 足够大时，$P\{|\xi|>M\}$ 便足够小.

这种随机变数在实践中有着广泛的应用. 例如：理论与实践都证明测量误差 ξ 服从正态 $N(0,\sigma)$，σ 的小或大，反映了密度函数的峰两旁的陡或平坦，亦即反映了测量的精度是高还是低. 又如射击对中心点的横向偏差与纵向偏差亦都分别服从正态分布. 其它，诸如机械制造过程中所发生的误差；人的身长；海洋波浪的高度；电子管中的噪声电流或电压；飞机材料的疲劳应力等等也都相当准确地服从这种"中间大，两头小"的正态分布.

上面提到的测量误差、人的身高、波浪高度、电子管噪声电流等以及上例出现的射击误差等等之所以能够相当近似地服从正态分布，就是因为它们都有一个共同的特点：它们都可以看作许多微小的，独立的随机因素作用的总后果，而每一个因素的影响都很小. 例如，射击的误差就受着许多随机因素诸如空气的温湿度、风速、气压、射击仪器的随机抖动、瞄准者情绪的随机波动等等的综合影响. 在正常的情况下，每一种因素都不应起着压倒一切的主导作用. 根据后面第五章讲到的中心极限定理，便可知具有这种特点的随机变数，一般都可以认为近似地服从正态分布，这正是正态分布在理论与实践上都极其重要的原因.

下面我们指出，即使对于我们已熟知的二项分布与泊松分布，当试验次数无限增大时(在二项分布的情形)，或平均出现次数 λ 无限增大时(在泊松分布的情形)，它们也都是趋于正态分布的. 因此，这也提供了一种近似计算二项分布或泊松分布的途径.

定理 2.1.3 (棣莫佛尔-拉普拉斯)①

① Abraham DeMoivre, The Doctrine of Chance, 1718. Pierre S. Laplace, Théorie analytique des probabilités, 1812.

设某事件 A 出现的概率为常数 $p(0<p<1)$，在 n 次独立试验中事件 A 恰好出现 k 次的概率记为 $P_n(k)$，现任取一个区间 $[a,b]$，记

$$x_k=\frac{k-np}{\sqrt{np(1-p)}},\quad k=1,2,\cdots$$

则有：凡 k 使得 $a\leqslant x_k\leqslant b$，必成立着

$$\lim_{n\to\infty}\sqrt{npq}P_n(k)\Big/\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_k^2}=1,\qquad(2.1.21)$$

其中 $q=1-p$. 并且，这个收敛在区间 $[a,b]$ 上是一致的.

这个定理的证明待学完第五章后由读者自已去完成. 这里举例说明它的应用.

例 2.1.19 设生产某种产品，出现废品的概率已知为 0.005，问任取 10000 件产品，其中恰好有 40 件废品的概率是多少？

解 由(2.1.21)可知

$$P_n(k)\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2}.\qquad(2.1.22)$$

由题设，$n=10000$，$p=0.005$，$q=1-p=0.995$，$k=40$. 故

$$\sqrt{npq}=\sqrt{49.75}\approx7.05,\quad \frac{k-np}{\sqrt{npq}}\approx-1.42;$$

$$P_{10000}(40)\approx\frac{1}{7.05\sqrt{2\pi}}e^{-\frac{1.42^2}{2}}\approx\frac{0.1456}{7.05}\approx0.0206.$$

$\left(\text{这里}\frac{1}{\sqrt{2\pi}}e^{-\frac{1.42^2}{2}}\approx0.1456\text{是查}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\text{的表求得的.}\right)$

为了应用的方便，将上面的定理推演成另一种形式是有好处的. 证明见定理 5.4.2.

定理 2.1.4 (隶莫佛尔-拉普拉斯积分极限定理)

若随机变数 ξ_n 服从二项分布 $b(n,p)$，$0<p<1$，即 $P(\xi_n=k)=C_n^kp^k(1-p)^{n-k}$，$(k=0,1,\cdots,n)$ 则

$$\lim_{n\to\infty}P\left(a\leqslant\frac{\xi_n-np}{\sqrt{npq}}\leqslant b\right)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.\qquad(2.1.23)$$

例 2.1.20 在上例中求废品数不大于 70 的概率.

解 设 n 件产品中出现的废品数为 ξ_n，则由 $np=50$，得

$$P(\xi_n \leqslant 70) = P\left(\frac{-50}{\sqrt{49.75}} \leqslant \frac{\xi_n - np}{\sqrt{npq}} \leqslant \frac{20}{\sqrt{49.75}}\right)$$

$$= P\left(-7.09 \leqslant \frac{\xi_n - np}{\sqrt{npq}} \leqslant 2.84\right) \approx \frac{1}{\sqrt{2\pi}} \int_{7.09}^{2.84} e^{-\frac{x^2}{2}} dz$$

$$= \Phi(2.84) - \Phi(-7.09) = 0.9975.$$

这里，在 $x \geqslant 0$ 的情形，$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$ 有表可查(见附录). 在 $x < 0$ 的情形，利用对称性 $\Phi(x) = 1 - \Phi(-x)$，仍可运用查表以求出 $\Phi(x)$.

例 2.1.21 随机抛硬币 100 次，设出现正面的概率为 0.5，问"正面出现次数小于 60、大于 50"的概率是多少？

解 记出现正面的次数为 ξ_n，现 $n=100$，$p=0.5$，$np=50$，$\sqrt{npq} = \sqrt{25} = 5$，故由(2.1.23)

$$P(50 < \xi_{100} < 60) = P\left(\frac{50-50}{5} < \frac{\xi_{100}-50}{5} < \frac{60-50}{5}\right)$$

$$= P\left(0 < \frac{\xi_{100}-50}{5} < 2\right) \approx \frac{1}{\sqrt{2\pi}} \int_{2}^{2} e^{-\frac{x^2}{2}} dx = 0.4772.$$

积分极限定理 2.1.4 还有许多应用. 这里再举一个例子.

例 2.1.22 (座位问题) 设由甲地至乙地有两种交通工具(例如，可以坐汽车直达，比较快但路费稍多些，也可走水陆联运，时间稍长路费却稍低)，每个旅客以 $\frac{1}{2}$ 的概率选择一种交通工具，假设有 1000 个旅客同时每天由甲地出发至乙地，若要求在 100 次中有 99 次有足够座位，(从各种交通工具的管理人员的角度看)问各种交通工具应各设多少座位？

解 1000 个旅客到来，可看作是作了 1000 次贝努里试验，对

于交通工具 A 来说，每次被使用的概率是 0.5，若有 s 个座位，则旅客座位够坐时的概率是 $\sum_{k=0}^{s} P_{1000}(k) \approx \Phi\left(\frac{s-np}{\sqrt{npq}}\right)$.

按要求 $\sum_{k=0}^{s} P_{1000}(k) \geqslant 0.99$，亦即要求

$$1-\Phi\left(\frac{s-500}{\sqrt{250}}\right) \leqslant 0.01.$$

查表可得

$$\frac{s-500}{\sqrt{250}} \geqslant 2.33.$$

亦即 $s \geqslant 536.84$.

故对交通工具 A 至少需设 537 个座位，则在 100 次中每次有 1000 个旅客随机选取交通工具时，最多只有 1 次发生座位不够. 对另一交通工具 B 也可作同样讨论.

下面的定理说明泊松分布的正态逼近.

*定理 2.1.5 对任意的 $a<b$，有

$$\lim_{\lambda \to \infty} \sum_{\alpha<k<\beta} P(k;\ \lambda) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \qquad (2.1.24)$$

其中 $\begin{cases} \alpha = \lambda + a\lambda^{\frac{1}{2}}, \\ \beta = \lambda + b\lambda^{\frac{1}{2}} \end{cases} \left(或 \begin{cases} a = \frac{\alpha-\lambda}{\sqrt{\lambda}}, \\ b = \frac{\beta-\lambda}{\sqrt{\lambda}} \end{cases} \right)$.

这一定理的证明留待第五章里读者自己去完成.

例 2.1.23（线路问题）设电话交换台 A 要为 2000 个用户服务，在最忙时，平均每小时每户打电话占线两分钟，(亦即每户用线的概率可设为 $\frac{2}{60}=\frac{1}{30}$)，假设各户打电话是相互独立的，问若想以 99% 的可能性满足用户的要求，最少需设多少条线路？

解　呼叫次数可看作服从泊松分布，由泊松近似中讲过$\lambda=np=2000\cdot\frac{1}{30}\approx66.67$. 现要求最少的线路 N，使

$$0.99<\sum_{k=0}^{N}P(k,66.67)\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{N-66.67}{\sqrt{66.67}}}e^{-\frac{x^2}{2}}dx.$$

查表可得，上式右端等于 0.99 时

$$\frac{N-66.67}{\sqrt{66.67}}=2.327,$$

故

$$N\geqslant66.67+(2.327)(8.027)\approx85.3,$$

亦即，安装86条线路就可以了.

二项分布由泊松定理(定理 2.1.1)可以用泊松分布近似：若

$$\lim_{n\to\infty}np_n=\lambda>0,$$

则当 n 足够大时，有

$$b(k;n,p_n)\approx P(k;np_n).$$

由这里的棣莫佛尔-拉普拉斯定理(定理 2.1.3)又可以用正态分布近似：若 $0<p<1$，则当 n 够大时，有

$$b(k;n,p)\approx\frac{1}{\sqrt{2\pi np(1-p)}}e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{np(1-p)}}\right)^2}.$$

自然要问，两种近似各在何种情况下更为适合？对照三种分布的数值表，可以发现，正如以前讲过的，当 p 很小的时候，即使 n 不用很大，用泊松分布近似二项分布，已经符合得相当地好(见例 2.1.11 前面一表的说明)．但是，在这种情形下，用正态分布去近似二项分布，却会产生较大的误差(直观上也可想像得到，p 很小，n 又不大，那末 $np=\lambda$ 一定不会很大；根据这里的定理 2.1.5，正态分布就不能很好地近似泊松分布，因而也就不能近似被泊松分布十分逼近的二项分布).

在 p 既不接近于 0 也不接近于 1 的情形(当 p 接近于 1 时，则 $q=1-p$ 便接近于 0)，用正态分布去近似二项分布，效果就较好.

下表是用泊松分布与正态分布去近似二项分布的比较.

k	$b(k;2500,0.02)$	$N(50,7)$	$P(k;50)$
25	.0000	.0001	.0000
30	.0006	.0010	.0007
35	.0052	.0057	.0054
40	.0212	.0205	.0215
45	.0460	.0442	.0458
50	.0569	.0570	.0563
55	.0424	.0442	.0422
60	.0199	.0206	.0201
65	.0061	.0057	.0063
70	.0013	.0010	.0014
75	.0002	.0001	.0002

例 2.1.24（韦布分布）若随机变数 ξ 的密度函数由下式给出

$$f(x)=\begin{cases}\dfrac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}, & \text{当 } x>\nu,\\ 0, & \text{当 } x\leqslant\nu,\end{cases} \tag{2.1.25}$$

其中，$m>0$ 称为形状参数，ν 称为位置参数，$x_0>0$ 称为尺度参数，则称 ξ 服从韦布分布.

参数 m,ν,x_0 的几何意义的示意图如图 2.1.13.

可以证明，上面给出的 $f(x)$ 满足(2.1.17)式.

显然，$f(x)\geqslant 0$，现在来证 $\int_{-\infty}^{\infty}f(x)dx=1$.

事实上

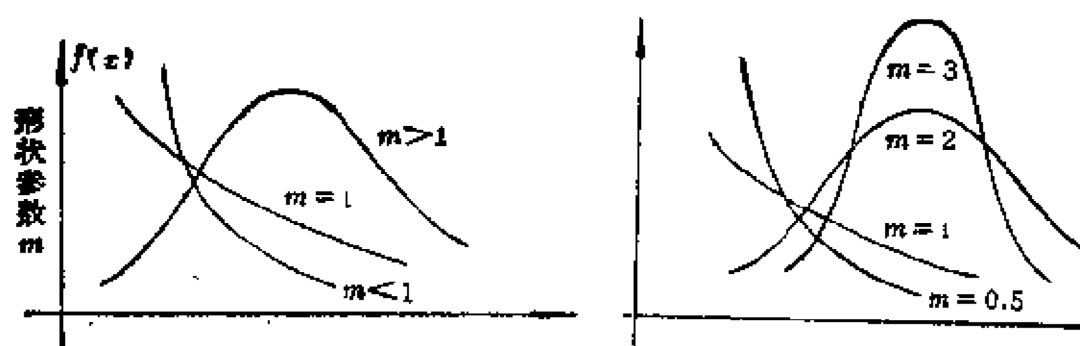
$$\int_{-\infty}^{\infty}f(x)dx=\frac{m}{x_0}\int_{\nu}^{\infty}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}dx,$$

令 $u=\dfrac{(x-\nu)^m}{x_0}$，上式右端变为

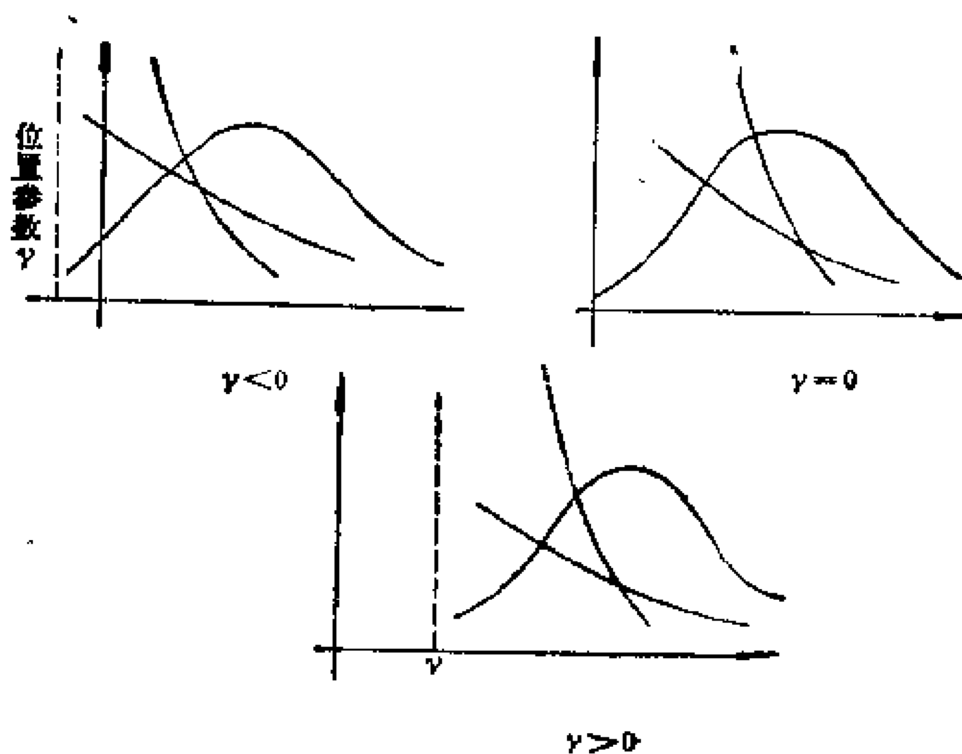
$$\int_0^{\infty}e^{-u}du=1.$$

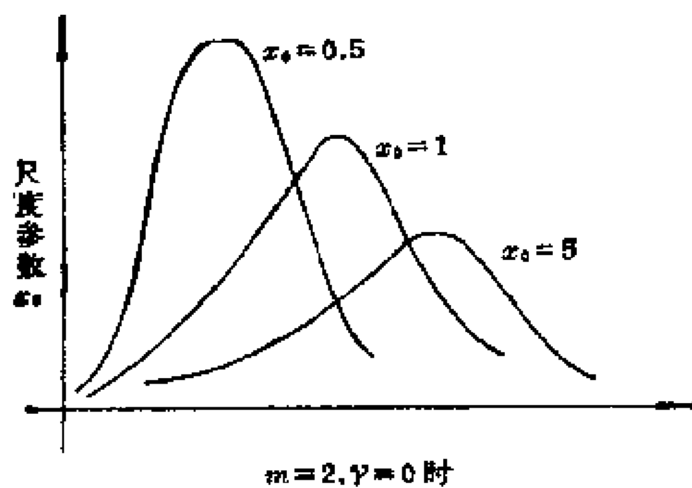
下面说明韦布分布的物理模型.

设有一条由 n 个同类型的环构成的链，两端受大小相等，方



$\gamma=0, x_0=1$ 时，m 值的影响





$m=2, \gamma=0$ 时

图 2.1.3　韦布分布参数对密度函数的影响

向相反的力 x 之作用如图

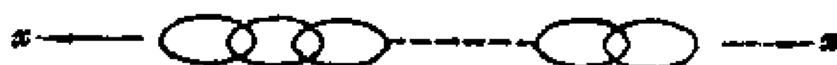


图 2.1.14

以 ξ 表单个环不被拉断所能承受(最大)的力，则 ξ 是一个随机变数．显然，只要一个环被力 x 拉断，则整条链便失效．故链失效的概率为 $P(\xi<x)$．现求 $P(\xi<x)$ 的表达式．

注意到单个环不被拉断的概率是 $1-P(\xi<x)$．令

$$1-P(\xi<x)=e^{-g(x)},$$

若环的最低(最大)承受力为 ν (常数)，则拉力 x 小于 ν 时，环不断．故

$$g(x)=\begin{cases}0, & \text{当 } x\leqslant\nu,\\ \geqslant 0, & \text{当 } x>\nu.\end{cases}$$

考虑到正幂函数是一种最基本的函数，并且满足 x 愈大时，$g(x)$ 愈大(力愈大，拉断的概率愈大)的要求，故不妨取

$$g(x)=\begin{cases}\dfrac{(x-\nu)^m}{x_0}, & \text{当 } x>\nu,\\ 0, & \text{当 } x\leqslant\nu,\end{cases}$$

其中 m,x_0 为常数且 $x_0>0,m>0$．于是得

$$P(\xi<x)=1-e^{-g(x)}=\begin{cases}1-e^{-\frac{(x-\nu)^m}{x_0}}, & \text{当 } x>\nu,\\ 0, & \text{当 } x\leqslant\nu.\end{cases}$$

若记 $F(x)=P(\xi<)$，则 x $f(x)=F'(x)$．即

$$f(x)=F'(x)=\begin{cases}\dfrac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}, & \text{当 } x>\nu,\\ 0, & \text{当 } x\leqslant\nu.\end{cases}$$

(2. 1. 26)

由微积分学中的牛顿-莱伯尼兹公式，知

$$P(a \leqslant \xi < b) = F(b) - F(a) = \int_a^b F'(x)\,dx = \int_a^b f(x)\,dx.$$

由此可见函数 (2.1.26) 确是给出单个环的强度 ξ 所服从的密度函数.

以上说明,当计算链的失效概率(或单个环被拉断的概率) 这种物理模型时,便需用到韦布分布.

显然,这一模型适用于:由某一局部失效,便引起全部机能停止的现象.(例如:交流电子管收音机中某一电子管的失效,便会使整部收音机不能收音,如假设各电子管的各个参数大致一样,便可应用韦布分布去研究处理.)

韦布分布最有用的特殊情况之一是指数分布,它的密度函数是

$$f(x) = \begin{cases} \lambda e^{-\lambda(x-\nu)}, & \text{当 } x > \nu, \\ 0, & \text{当 } x \leqslant \nu. \end{cases} \tag{2.1.27}$$

这只需在(2.1.25)中取参数 $m=1$, $\frac{1}{x_0}=\lambda$,便得.

例 2.1.25(Γ-分布) 若随机变数 ξ 的密度函数由下式给出

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} x^{\alpha} e^{-\frac{x}{\beta}}, & x > 0, \\ 0, & x \leqslant 0. \end{cases} \tag{2.1.28}$$

其中 $\alpha > -1$, $\beta > 0$,则称 ξ 服从 Γ-分布, 或写为 ξ 服从 $\Gamma[\alpha,\beta]$ (Γ-分布有时也称为皮尔逊 III-型分布.)

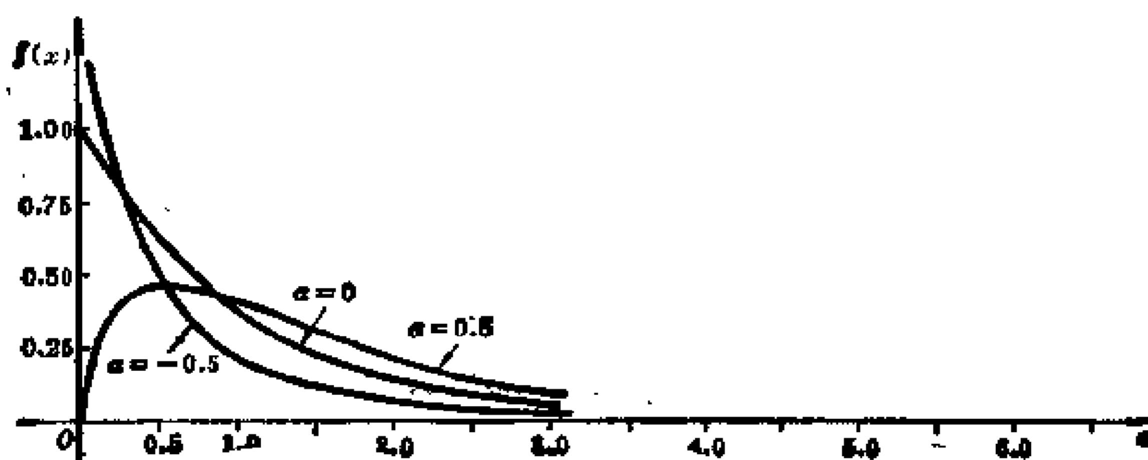
参数 α,β 的几何意义,如图 2.1.15 及图 2.1.16.

这里

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx = 2\int_0^{\infty} y^{2\alpha-1} e^{-y^2}\,dy \quad (\alpha > 0), \tag{2.1.29}$$

是微积分中所熟知的 Γ-函数,它有下列公式

$$\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1); \tag{2.1.30}$$



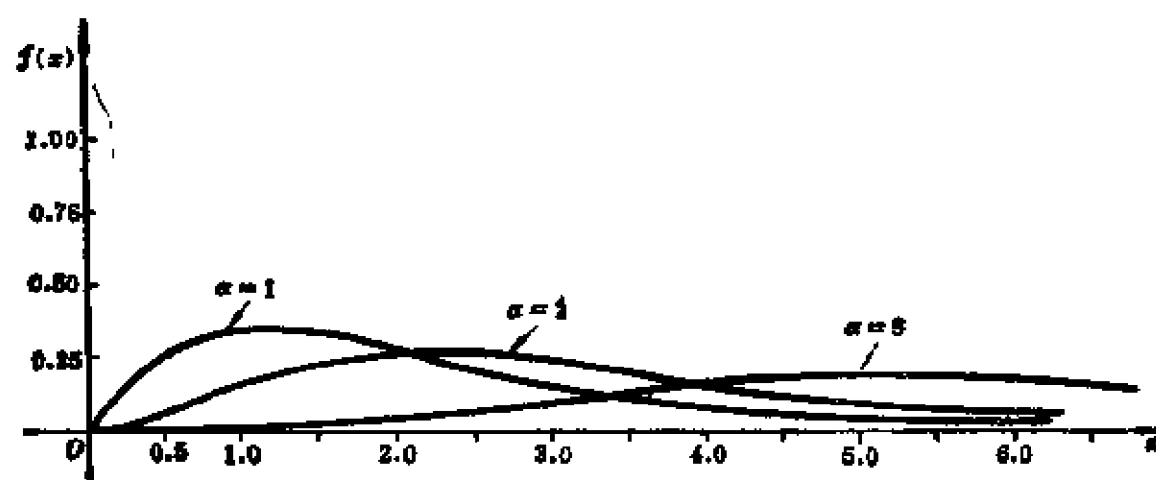


图 2.1.15　$\beta=1$，α 值的影响

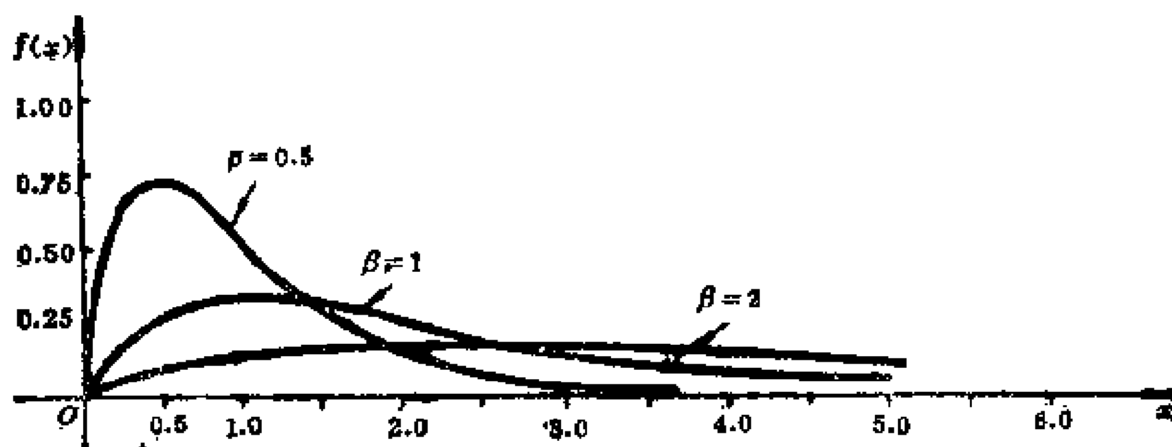


图 2.1.16　$\alpha=1$ 时，β 值的影响

$$\Gamma(1)=1;\qquad \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}.$$

需要验证上面给出的 $f(x)$ 满足(2.1.17)式.

显然，$f(x)\geqslant 0$，下面来证 $\int_{-\infty}^{\infty} f(x)\,dx=1$.

事实上

$$\int_0^{\infty}\frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}dx=\frac{1}{\Gamma(\alpha+1)}\int_0^{\infty}y^{\alpha}e^{-y}dy \qquad \left(令\ y=\frac{x}{\beta}\right)$$

$$=\frac{1}{\Gamma(\alpha+1)}\cdot\Gamma(\alpha+1)=1.$$

顺便指出，当 $\alpha=0$，我们重新由 Γ-分布密度函数得出了指数分布的密度函数：

$$f(x)=\begin{cases}\frac{1}{\beta}e^{-\frac{x}{\beta}}, & 0\leqslant x<\infty;\\ 0, & 其它.\end{cases}$$

Γ-分布既在推导统计学中有重要地位的 χ^2-分布、t-分布和 F-分布时很有用，又在水文统计，或最大风速，最大风压的概率计算中经常用到. 它是一种重要的非正态分布.

例 2.1.26(对数正态分布) 设随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}\frac{\lg e}{\sqrt{2\pi}\,\sigma x}e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2}, & 当\ x>0;\\ 0, & 当\ x\leqslant 0,\end{cases} \qquad (2.1.31)$$

其中 $\sigma>0$，μ 为常数，则称 ξ 服从对数正态分布.

参数的几何意义，我们示意如图 2.1.17.

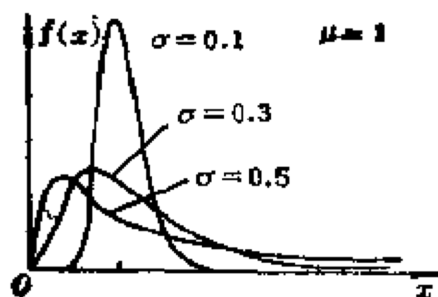


图 2.1.17

可以证明，$f(x)$ 满足(2.1.17). 显然，$f(x)\geqslant 0$. 现证

$$\int_{-\infty}^{\infty} f(x)\,dx=1.$$

事实上

$$\int_0^{\infty}\frac{\lg e}{\sqrt{2\pi}\,\sigma x}e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2}dx$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy=1\qquad\left(令\ y=\frac{\lg x-\mu}{\sigma}\right).$$

对数正态分布在寿命试验中经常都会遇到．它的物理模型如下：

设某一偶然现象由一连串相互独立的正的偶然因素作用引起，若前者记为ξ，后者为记$\xi_n(n=1,2,\cdots)$，且它们的关系为

$$\xi=\xi_1\cdot\xi_2\cdots\cdots\xi_n\cdots,$$

取对数，便得

$$\lg\xi=\sum_n\lg\xi_n.$$

若所有的$\lg\xi_n$也都很微小且相互独立，于是由以后叙述的中心极限定理，可知$\lg\xi$服从正态分布，设服从$N(\mu,\sigma)$．这时，ξ便服从上面的对数正态分布．事实上

$$F_\xi(x)=P(\xi<x)=P(\lg\xi<\lg x)$$

$$=\int_{-\infty}^{\lg x}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,$$

令

$$f_\xi(x)=\frac{dF_\xi(x)}{dx}=\frac{\lg e}{\sqrt{2\pi}\,\sigma x}e^{-\frac{(\lg x-\mu)^2}{2\sigma^2}},\quad x>0.$$

由假定$\xi>0$，故$F_\xi(x)=P(\xi<x)=0$（当$x\leqslant0$），从而

$$f_\xi(x)=\frac{dF_\xi(x)}{dx}=0\qquad(当x\leqslant0).$$

上面介绍了几种常用的离散型与连续型随机变数的分布列与密度函数．其中离散型的二项分布，泊松分布与连续型的正态分布最为常用，必须加以特别的注意．

三、分布函数及其基本性质

由上面所述，离散型随机变数，其取值的概率分布情况，可用

分布列来描述．对于连续型随机变数，其取值的概率分布情况，则用密度函数的积分来描述．但是除了这两种随机变数外，还有连续取值而非连续型的(即密度函数不存在)或混合型的．为了理论研究的方便，必须给出一个描述随机变数取值的概率分布情况的统一的方法．一个常用且较为简单的方法是引入如下定义：

定义 2.1.4(分布函数)　设 ξ 为一随机变数．对任意 $x\in R_1$，令

$$F(x)=P\{\xi<x\}①.$$

称 $F(x)$ 为 ξ 的分布函数.

由定义 2.1.4 立即可以得到：当 $a<b$ 时

$$P\{a\leqslant\xi<b\}=F(b)-F(a). \tag{2.1.32}$$

事实上，因为

$$\{a\leqslant\xi<b\}=\{\xi<b\}-\{\xi<a\},$$

且

$$\{\xi<a\}\subset\{\xi<b\},$$

故

$$P\{a\leqslant\xi<b\}=P\{\xi<b\}-P\{\xi<a\}=F(b)-F(a).$$

对于离散型随机变数，其分布函数为

$$F(x)=\sum_{x_k<x}P\{\xi=x_k\},$$

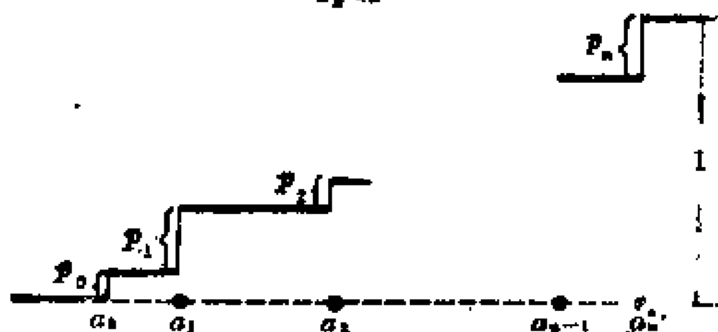


图 2.1.18　离散型随机变数的分布函数

① $\{\xi<x\}$ 有时也写为 $\{\xi\in(-\infty,x)\}$.

其中求和是对所有满足不等式 $x_k<x$ 的指标 k 进行的．它的示意图如图 2.1.18

对于连续型随机变数，其分布函数为

$$F(x)=\int_{-\infty}^{x} f(x)dx, \qquad x\in R_1.$$

示意图如图 2.1.19.

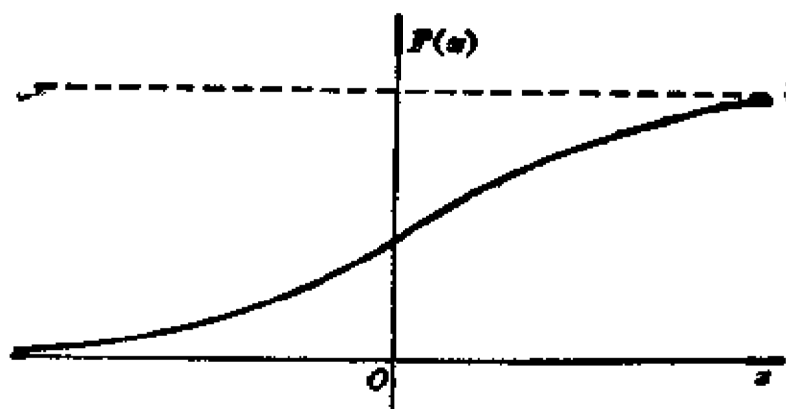


图 2.1.19 连续型随机变数的分布函数

若 $f(x)$ 连续，则

$$\frac{d}{dx}F(x)=f(x).$$

反过来，对于随机变数 ξ，若知道了它的分布函数 $F(x)$:

$$F(x)=P(\xi<x)$$

则当 ξ 为离散型随机变数时，它的分布列也可由分布函数所确定．事实上，

$$P_n=P(\xi=x_n)=\lim_{k\to\infty}P\left(x_n\leqslant\xi<x_n+\frac{1}{k}\right)$$

$$=\lim_{k\to\infty}\left[P\left\{\xi<x_n+\frac{1}{k}\right\}-P\{\xi<x_n\}\right]$$

$$=\lim_{k\to\infty}\left[F\left(x_n+\frac{1}{k}\right)-F(x_n)\right].$$

故 $P_n(n=1,2,\cdots)$ 可由分布函数求出．

当 ξ 为连续型随机变数时，它的密度函数设为 $f(x)$，则由(2.1.32)式有

$$\int_a^b f(x)dx = P(a \leqslant \xi < b) = F(b) - F(a).$$

故连续型随机变数取值之概率分布情况亦可由分布函数所唯一确定.

下述定理给出了分布函数的基本性质.

定理 2.1.6 设 $F(x)$ 为随机变数 ξ 的分布函数，则

1° $F(x)$ 为单调不降；

2° $F(x)$ 为左连续；

3° $\lim\limits_{x\to-\infty} F(x)=0$, $\quad \lim\limits_{x\to+\infty} F(x)=1$.

(这两个式子，也常常记为 $F(-\infty)=0, F(\infty)=1$.)

证 1° 设 $x_1<x_2$，由 $\{\xi<x_1\}\subset\{\xi<x_2\}$ (即左边事件发生，右边必发生)得

$$F(x_1)=P(\xi<x_1)\leqslant P(\xi<x_2)=F(x_2),$$

因而证明了 $F(x)$ 单调不降.

2° 对任意 $x\in R_1$，由于 $F(x)$ 单调不降，要证 $F(x)$ 左连续性，只需证：$\lim\limits_{n\to\infty}F\left(x-\frac{1}{n}\right)=F(x)$.

事实上

$$F\left(x-\frac{1}{n}\right)=P\left(\xi<x-\frac{1}{n}\right),$$

由于 $\bigcup\limits_{n=1}^{\infty}\left\{\xi<x-\frac{1}{n}\right\}=\{\xi<x\}$，且 $\left\{\xi<x-\frac{1}{n}\right\}\subset\left\{\xi<x-\frac{1}{n+1}\right\}$，故利用概率的连续性定理，可得

$$\lim_{n\to\infty}F\left(x-\frac{1}{n}\right)=\lim_{n\to\infty}P\left(\xi<x-\frac{1}{n}\right)=P(\xi<x)=F(x).$$

这就证明了 $F(x)$ 的左连续性.

3° 由于 $\bigcap_{n=1}^{\infty}\{\xi<-n\}=\varnothing$，且 $\{\xi<-n\}\supset\{\xi<-(n+1)\}$，故知

$$\lim_{n\to\infty}F(-n)=\lim_{n\to\infty}P(\xi<-n)=P(\varnothing)=0.$$

再由于 $\bigcup_{n=1}^{\infty}\{\xi<n\}=\Omega$（必然事件），且 $\{\xi<n\}\subset\{\xi<n+1\}$，故知

$$\lim_{n\to\infty}F(n)=\lim_{n\to\infty}P(\xi<n)=P(\Omega)=1.$$

读者可试作出以前学过的服从各种分布的随机变数的分布函数的图形，从而进一步验证上述性质的真确.

值得注意的是，定理 2.1.6 的逆定理也成立：

若实值函数 $F(x)$，$x\in R_1$，满足定理 2.1.2 中的 1°—3°，则必存在一个概率空间上的随机变数 ξ，以 $F(x)$ 为其分布函数．因此 1°—3° 完全刻划了一个随机变数的分布函数．此逆定理的证明需用到测度论知识，所以从略（可参看[1] § 2.2）.

最后，在本节之末，让我们指出既不是离散型，也不是连续型的随机变数之存在性.

例 2.1.27 设

$$F(x)=\begin{cases}0, & \text{当 } x\leqslant 0;\\ \dfrac{1+x}{2}, & \text{当 } 0<x\leqslant 1;\\ 1, & \text{当 } x>1.\end{cases}$$

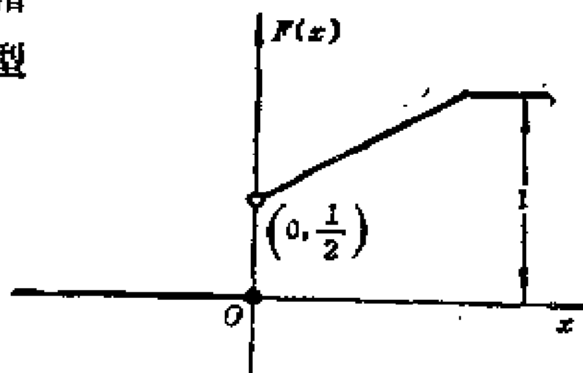


图 2.1.20 非离散有间断的分布函数

如图 2.1.20.

显然 $F(x)$ 满足定理 2.1.2 之 1°—3°，故是分布函数，但既非离散型（$F(x)$ 只有一个跳跃点，但又不是阶梯形跳跃），也非连续型（$F(x)$ 不连续，故更不可能表为非负函数的上限变动的积分）.

上述例子是分布函数有间断点因而是非连续型的情形．自然会问，有没有分布函数为连续，但又不是绝对连续（即不能表为非负函数变动上限的积

分)的情形，因而是既非连续型，又非离散型的例子；学过实变函数的读者知道，若 $F(x)$ 为 $x\in[0,1]$ 上的康托函数①，则 $F(x)$ 是连续上升的，且满足 $F(0)=0, F(1)=1$. 让我们将 $F(x)$ 作连续的常数延拓至 $(-\infty, 0)$ 与 $(0,\infty)$ 上. 亦即，令 $F(x)=0$，当 $x\leqslant 0$；及 $F(x)=1$，当 $x\geqslant 1$，则延拓后所得的 $F(x)$ 就是一个连续的分布函数但又不是绝对连续的. 因此它既非连续型，也非离散型.

§2.2 多维随机变数及其分布函数

一、二维分布函数及其基本性质

在生产实际与理论研究中，都常常会遇到这种情况：需要同时用几个随机变数才能较好地描绘某一试验或现象. 例如，炮弹在地面的命中点的位置是由一对随机变数(两个坐标)来确定；电子放大器的干扰电流由其振幅和相位这两个随机变数来给定；飞机(的重心)在空中的位置由三个随机变数(三个坐标)来确定，等等. 我们称 n 个随机变数 $\xi_1,\xi_2,\cdots,\xi_n$ 的总体 $\xi=(\xi_1,\xi_2,\cdots,\xi_n)$ 为 n 元随机变数(或 n 维随机变数). 由于二维与 n 维没有什么原则的区别，故为简单及容易理解起见，下面着重讨论二维情形. ($n=1$ 时的一维随机变数，就是 §2.1 所讨论的随机变数)与 §2.1 相仿，引入如下定义：

定义 2.2.1 (二维联合分布函数) 设 ξ,η 为定义在同一个概率空间 $(\Omega,\mathscr{F},P)$ 上的两个随机变数②，则 (ξ,η) 称为二维随机变数. 对任意 $x,y\in R_1$

令

$$F(x,y)=P\{\omega:\xi(\omega)<x,\eta(\omega)<y\}, \qquad (2.2.1)$$

或简写为

① 见 И. П. 那汤松：《实变函数论》上册(徐瑞云译，商务印书馆 1953) p.p. 263—264.

② 今后无特别声明，当讨论随机变数时，均是定义于同一概率空间上，且有时称 (ξ,η) 为随机向量.

$$F(x,y)=P\{\xi<x,\eta<y\}.$$

称 $F(x,y)$ 为 (ξ,η) 的联合分布函数，或简称二维分布函数.

定理 2.2.1 设 $F(x,y)$ 为随机变数 (ξ,η) 的联合分布函数，则①

1° $F(x,y)$ 分别对 x 和 y 单调不降，即

$F(x_2,y)\geqslant F(x_1,y)$，当 $x_2>x_1$，

$F(x,y_2)\geqslant F(x,y_1)$，当 $y_2>y_1$，

2° $F(x,y)$ 对每个变元左连续；

3° $\lim\limits_{x\to-\infty}F(x,y)=0$，$\lim\limits_{y\to-\infty}F(x,y)=0$，

$$\lim_{\substack{x\to\infty\\y\to\infty}}F(x,y)=1.$$

这三个式子，也常简记为 $F(-\infty,\ y)=0$，$F(x,\ -\infty)=0$，$F(\infty,\infty)=1$.

4° 对任意四个实数 $a_1\leqslant b_1$，$a_2\leqslant b_2$，有

$$F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)\geqslant 0.$$

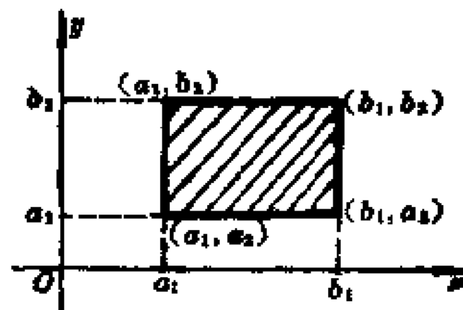


图 2.2.1

证 1°—3°之证明类似于定理2.1.6.仅需证 4°.如图 2.2.1，给出 $a_1\leqslant b_1,a_2\leqslant b_2$，显然

$$\begin{aligned}0&\leqslant P(a_1\leqslant\xi<b_1,a_2\leqslant\eta<b_2)\\&=P\{\xi<b_1,\eta<b_2\}-P\{\xi<a_1,\eta<b_2\}\\&\quad-P\{\xi<b_1,\eta<a_2\}+P\{\xi<a_1,\eta<a_2\}\\&=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2).\ \blacksquare\end{aligned}$$

同样可以指出，定理 2.2.1 之逆定理也是成立的.

若二元实值函数 $F(x,y)\ (x,y\in R_1)$ 满足定理2.2.1 的1°—4°，

① 容易证明，3°与4°蕴含了 1°

则必存在随机变数 ξ 与 η(在某一特定的概率空间上)，使 $F(x,y)$ 是 (ξ,η) 的联合分布函数.

这一逆定理之证明同样需用到测度论的知识，所以从略.

值得注意的是：在多维随机变数时与一维随机变数时不一样，刻划一个联合分布函数需要加设条件 4°. 下面举一例说明条件4°不能从条件 1°—3° 推出.

图 2.2.2

例 2.2.1 令 $F(x,y)=\begin{cases}1, & 当\ x+y>-1,\\ 0, & 当\ x+y\leqslant -1.\end{cases}$

显然，$F(x,y)$ 满足 1°—3°，但不满足 4°，因为

$$F(1,1)-F(1,-1)-F(-1,1)+F(-1,-1)$$
$$=1-1-1+0=-1.$$

与一维时情形一样，二维分布也有最常用的两大类型.

1) 离散型. 当随机变数 ξ 与 η 只可能取有限个或可列个值时，称其联合分布函数为离散型分布函数，称二维随机变数 (ξ,η) 为离散型的.

对于二维离散型随机变数，与一维时情形一样，用表格给出其取值的概率分布情况，更为直观与方便：

表 2.2.1

ξ \ η	y_1	y_2	…	y_k	…
x_1	$p(1,1)$	$p(1,2)$	…	$p(1,k)$	…
x_2	$p(2,1)$	$p(2,2)$	…	$p(2,k)$	…
⋮	…	…	…	…	…
x_i	$p(i,1)$	$p(i,2)$	…	$p(i,k)$	…
⋮	…	…	…	…	…

其中 $p(i,k)=P(\xi=x_i,\eta=y_k)$，$p(i,k)\geqslant 0$，且

$$\sum_{i=1}^{\infty}\sum_{k=1}^{\infty}p(i,k)=1. \tag{2.2.2}$$

称 $p(i,k)\ (i,k=1,2,\cdots)$ 为概率函数，此时有

$$F(x,y)=\sum_{x_i<x}\ \sum_{y_k<y}P\{\xi=x_i,\eta=y_k\}=\sum_{x_i<x}\ \sum_{y_k<y}p(i,k).$$

例 2.2.2（二维两点分布） 设 ξ，η 由表 2.2.2 给出（其中 $0\leqslant p\leqslant 1$），则称 (ξ,η) 服从二维二点分布.（显然满足(2.2.2)式.）

表 2.2.2

ξ \ η	0	1
0	$1-p$	0
1	0	p

2）连续型. 若存在非负二元函数 $f(x,y)$，对任意 $x,y\in R_1$ 使

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv,$$

其中 $f(x,y)\geqslant 0$ 且满足

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1, \tag{2.2.3}$$

则 $F(x,y)$ 称为随机变数 (ξ,η) 的联合分布函数，称 (ξ,η) 为连续型的二维随机变数. 称其联合分布函数 $F(x,\ y)$ 为连续型的联合分布函数. 称 $f(x,y)$ 为 (ξ,η) 的联合密度函数.

如果联合密度函数 $f(x,y)$ 在点 (x,y) 的邻域内连续且存在 $g(x)$，满足

$$\int_{-\infty}^{\infty}g(x)dx$$

绝对收敛，使

$$|f(x,y)|<g(x),$$

则

$$\frac{\partial^2 F(x,y)}{\partial x \partial y}=f(x,y).$$

例 2.2.3 若二维随机变数(ξ,η)的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{(b_1-a_1)(b_2-a_2)}, & \text{当 } a_1\leqslant x\leqslant b_1,\ a_2\leqslant y\leqslant b_2;\\ 0, & \text{其它,}\end{cases}$$

则称(ξ,η)服从均匀分布.

可以证明,$f(x,y)$满足(2.2.3)式,$f(x,y)\geqslant 0$是显然的,且

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy=\int_{a_2}^{b_2}\int_{a_1}^{b_1}\frac{dxdy}{(b_1-a_1)(b_2-a_2)}=1.$$

例 2.2.4(正态分布) 设随机变数(ξ,η)的联合密度函数为:

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)(x-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]\right\}^{①},$$

其中$\sigma_1,\sigma_2>0$,m_1,m_2均为常数,$|r|<1$,则称(ξ,η)服从二维正态分布$N(m_1,\sigma_1;m_2,\sigma_2;r)$.

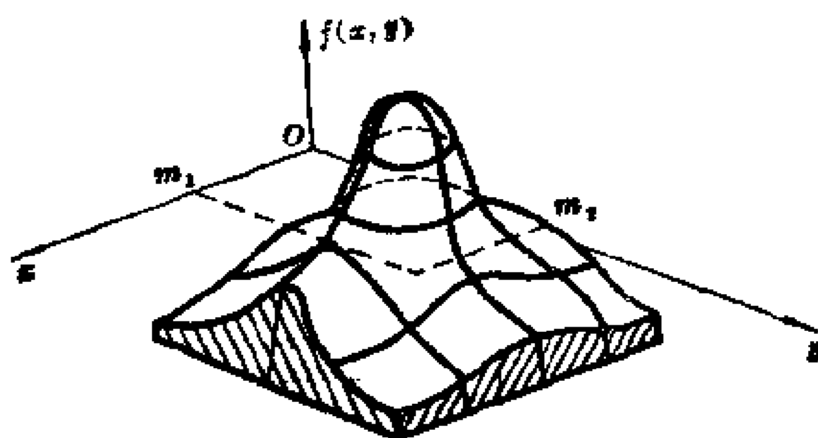


图 2.2.3 二维正态密度函数

① 此处记$\exp[\cdot]=e^{[\cdot]}$.

二维正态密度函数的示意图见图 2.2.3.

$f(x,y)\geqslant 0$ 是显然的，可以证明，$f(x,y)$ 满足(2.2.3)式. 即

证

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1.$$

为此，令 $\frac{x-m_1}{\sigma_1}=u$，$\frac{y-m_2}{\sigma_2}=v$，先计算 $\int_{-\infty}^{\infty} f(x,y)\,dy$，并以 $f_1(x)$ 表示，则

$$
\begin{aligned}
f_1(x)&=\int_{-\infty}^{\infty} f(x,y)\,dy\\
&=\frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}[u^2-2ruv+v^2]\right\}dv\\
&=\frac{1}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\exp\left\{-\frac{1}{2(1-r^2)}[(v-ru)^2\right.\\
&\qquad\qquad\left.+(1-r^2)u^2]\right\}dv\\
&=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{u^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\\
&\quad\cdot\exp\left\{-\frac{1}{2(1-r^2)}(v-ru)^2\right\}dv\\
&=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{u^2}{2}}=\frac{1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(x-m_1)^2}{2\sigma_1^2}},
\end{aligned}
\tag{2.2.4}
$$

故得

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=\int_{-\infty}^{\infty} f_1(x)\,dx=1.$$

若令

$$f_2(y)=\int_{-\infty}^{\infty} f(x,y)\,dx,$$

同样可得

$$f_2(y)=\frac{1}{\sigma_2\sqrt{2\pi}}e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}. \tag{2.2.5}$$

考虑

$$\left.\begin{aligned} F_1(x)&=\int_{-\infty}^{x} f_1(t)\,dt, \\ F_2(y)&=\int_{-\infty}^{y} f_2(t)\,dt. \end{aligned}\right\} \tag{2.2.6}$$

由(2.2.4)式及(2.2.5)式知，$F_1(x)$，$F_2(y)$是正态分布函数。注意到二维联合分布函数 $F(x,y)$的定义及(2.2.6)式，有

$$\begin{aligned} F_1(x)&=\int_{-\infty}^{x} f_1(t)\,dt=\int_{-\infty}^{x} dt\int_{-\infty}^{\infty} f(t,y)\,dy \\ &=\lim_{y\to\infty}\int_{-\infty}^{x}\int_{-\infty}^{y} f(t,v)\,dt\,dv=\lim_{y\to\infty}F(x,y) \\ &=F(x,\infty)=P\{\xi<x,\eta<\infty\} \\ &=P\{\xi<x\}①. \end{aligned}$$

同理

$$\begin{aligned} F_2(y)&=\int_{-\infty}^{y} f_2(t)\,dt=\int_{-\infty}^{y} dt\int_{-\infty}^{\infty} f(x,t)\,dx=F(\infty,y) \\ &=P\{\xi<\infty,\eta<y\}=P\{\eta<y\}. \end{aligned}$$

即

$$\left.\begin{aligned} F_1(x)&=F(x,\infty)=P\{\xi<x\}=F_\xi(x), \\ F_2(y)&=F(\infty,y)=P\{\eta<y\}=F_\eta(y). \end{aligned}\right\} \tag{2.2.7}$$

二、边沿分布

上面的(2.2.7)式说明，若二维联合分布函数，其中有一个变元让它趋于无穷，则其极限函数恰好是一维分布函数。对于这种特殊性质，我们将给它一个名称。为此，对任意两个随机变数，我们引入

定义 2.2.2(边沿分布) 设 $F(x,y)$为 ξ，η 的联合分布函数。令

---

① 这里采用定理 2.2.1 中 3°用过的类似记号，即记 $F(x,\infty)=\lim\limits_{y\to\infty}F(x,y)$及 $F(\infty,y)=\lim\limits_{x\to\infty}F(x,y)$。

$$\left.\begin{array}{l} F_1(x) = F(x,\infty), \\ F_2(y) = F(\infty, y), \end{array}\right\} \tag{2.2.8}$$

分别称 $F_1(x)$和 $F_2(y)$为 $F(x,y)$关于ξ和关于η的边沿分布函数，或简称为ξ和η的边沿分布函数.

根据定义 2.2.2,可知

$$\begin{aligned} F_1(x) &= F(x,\infty) = \lim_{y\to\infty} F(x,y) \\ &= P\{\xi<x, \eta<\infty\} = P\{\xi<x\} = F_\xi(x), \\ F_2(y) &= F(\infty, y) = \lim_{x\to\infty} F(x,y) \\ &= P\{\xi<\infty, \eta<y\} = P\{\eta<y\} = F_\eta(y). \end{aligned}$$

对于连续型随机变数(ξ,η),则有:

$$F_1(x) = F(x,\infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,v)\,du\,dv,$$

$$F_2(y) = F(\infty, y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f(u,v)\,du\,dv.$$

我们把

$$\left.\begin{array}{l} \int_{-\infty}^{\infty} f(x,y)\,dy, \\ \int_{-\infty}^{\infty} f(x,y)\,dx \end{array}\right\} \tag{2.2.9}$$

分别称为ξ,η的边沿密度函数，并分别以 $f_1(x)$及 $f_2(y)$表示. 这样有 $f_\xi(x)=f_1(x)$, $f_\eta(y)=f_2(y)$,其中 $f_\xi(x)$, $f_\eta(y)$分别为ξ,η的密度函数.

从(2.2.4)式和(2.2.5)式知，二维正态分布函数的边沿分布函数仍是正态分布函数.

对于具有概率函数$p(i,k)=P\{\xi=x_i,\eta=y_k\}$的二维离散型随机变数$(\xi,\eta)$,则有

$$
\left.\begin{aligned}
F_1(x)=F(x,\infty)=\sum_{x_i<x}\sum_k p(i,k),\\
F_2(y)=F(\infty,y)=\sum_{y_k<y}\sum_i p(i,k),
\end{aligned}\right\} \tag{2.2.10}
$$

符号$\sum\limits_{x_i<x}\sum\limits_k$表示对一切$x_i<x$及所有$y_k$的那些指标$i,k$求和。

同样，把

$$\sum_k p(i,k)$$

称为ξ的边沿概率函数，并以$p(i,\cdot)$表示．把

$$\sum_i p(i,k)$$

称为η的边沿概率函数，并以$p(\cdot,k)$表示．也有

$$
\left.\begin{aligned}
P_\xi(x_i)=p(i,\cdot)\quad(i=1,2,\cdots),\\
P_\eta(y_k)=p(\cdot,k)\quad(k=1,2,\cdots),
\end{aligned}\right\} \tag{2.2.10'}
$$

其中$P_\xi(x_i)$，$(i=1,2,\cdots)$；$P_\eta(y_k)$，$(k=1,2,\cdots)$分别为ξ，η的概率函数．

例 2.2.5　考虑随机变数ξ取值为$1,2,\cdots,n$，η取值为$1,2,\cdots,n$的二维随机变数(ξ,η)，其概率函数和边沿概率函数如表2.2.3

在例2.2.5中，当$n=2$时，可考虑作为具有等可能的随机抛掷两个硬币或随机抽查两个出生婴儿性别等的数学模型．表中表明

$$p(i,k)=\frac{1}{n^2},$$

ξ的概率函数

$$P_\xi(i)=p(i,\cdot)=\sum_k p(i,k)=\sum_{k=1}^{n}\frac{1}{n^2}=\frac{1}{n}$$

$$(i=1,2,\cdots,n).$$

同理，η的概率函数

$$P_\eta(k)=p(\cdot,k)=\frac{1}{n} \qquad (k=1,2,\cdots,n).$$

表 2.2.3

η \ ξ	1	2	$\cdots$	n	$p(\cdot,k)$
1	$\frac{1}{n^2}$	$\frac{1}{n^2}$	$\cdots$	$\frac{1}{n^2}$	$\frac{1}{n}$
2	$\frac{1}{n^2}$	$\frac{1}{n^2}$	$\cdots$	$\frac{1}{n^2}$	$\frac{1}{n}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
n	$\frac{1}{n^2}$	$\frac{1}{n^2}$	$\cdots$	$\frac{1}{n^2}$	$\frac{1}{n}$
$p(i,\cdot)$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$	$\sum_i\sum_k p(i,k)=1$

例 2.2.6 设 ξ 取值 0,1,2, η 取值为 0, 1 的二维随机变数 (ξ,η) 的概率函数如表 2.2.4:

$$p(0,j)=\frac{1}{4} \qquad (j=0,1), \qquad p(1,0)=\frac{1}{6},$$

$$p(1,1)=\frac{1}{8}, \qquad p(2,0)=\frac{1}{8}, \qquad p(2,1)=\frac{1}{12}.$$

表 2.2.4

η \ ξ	0	1	2	$p(\cdot,k)$
0	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{13}{24}$
1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{12}$	$\frac{11}{24}$
$p(i,\cdot)$	$\frac{1}{2}$	$\frac{7}{24}$	$\frac{5}{24}$	$\sum_i\sum_k p(i,k)=1$

我们有

$$P_\xi(i)=p(i,\cdot)=\begin{cases}\frac{1}{2} & i=0;\\ \frac{7}{24} & i=1;\\ \frac{5}{24} & i=2,\end{cases}$$

$$P_\eta(k)=p(\cdot,k)=\begin{cases}\frac{13}{24} & k=0;\\ \frac{11}{24} & k=1.\end{cases}$$

关于联合分布的边沿分布的概念，可推广到任意有限多维的情形；设$(\xi_1,\xi_2,\cdots,\xi_n)$的联合分布函数为$F(x_1,x_2,\cdots,x_n)$，令

$$F_{1,2,\cdots,k}(x_1,x_2,\cdots,x_k)=F(x_1,x_2,\cdots,x_k,\infty,\cdots,\infty)$$
$$=\lim_{x_{k+1},\cdots,x_n\to\infty}F(x_1,x_2,\cdots,x_k,\cdots,x_n)\quad(1\leqslant k\leqslant n).\tag{2.2.11}$$

称$F_{1,2,\cdots,k}(x_1,x_2,\cdots x_k)$为$(\xi_1,\xi_2,\cdots,\xi_k)$的边沿分布，同样可以证明它是$(\xi_1,\ \xi_2,\ \cdots,\ \xi_k)$的联合分布函数．上述的讨论同样对于$(\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_k})$也成立．

最后我们指出，如同一维正态分布一样，二维或多维正态分布在许多实际问题中也起着重要的作用．如射击问题中，炮弹在平面上的弹着点就可用服从二维正态分布的二维随机变数来刻划它．

§2.3 相互独立随机变数，条件分布

一、相互独立随机变数

我们在第一章引入了事件独立性的概念，现在把这概念引用到随机变数上来．

定义 2.3.1 设ξ,η为两个随机变数，若对任意实数x,y，有

$$P\{\xi<x,\eta<y\}=P(\{\xi<x\}\cap\{\eta<y\})$$
$$=P\{\xi<x\}\cdot P\{\eta<y\},\tag{2.3.1}$$

则称 ξ,η 相互独立.

它的意义是事件 $\{\xi<x\}$ 与事件 $\{\eta<y\}$ 相互独立.

显然,若 $F(x, y)$ 为 (ξ, η) 的联合分布函数, 则(2.3.1)式等价于

$$F(x,y)=F_1(x)F_2(y), \tag{2.3.1$'$}$$

其中 $F_1(x),F_2(y)$ 分别为 ξ,η 的分布函数.

对于离散型随机变数, (2.3.1)$'$ 等价于: 对任意实数 x_i,y_k, 有①

$$P\{\xi=x_i,\eta=y_k\}=P\{\xi=x_i\}\cdot P\{\eta=y_k\}, \tag{2.3.2}$$

亦即

$$p(i,j)=p(i,\cdot)p(\cdot,j).$$

事实上,若(2.3.2)式成立,则有

$$\begin{aligned}
F(x,y)&=\sum_{x_i<x}\sum_{y_k<y}P\{\xi=x_i,\eta=y_k\}\\
&=\sum_{x_i<x}\sum_{y_k<y}P\{\xi=x_i\}P\{\eta=y_k\}\\
&=\left[\sum_{x_i<x}P\{\xi=x_i\}\right]\cdot\left[\sum_{y_k<y}P\{\eta=y_k\}\right]\\
&=F_1(x)F_2(y),
\end{aligned}$$

即(2.3.1)$'$ 成立.

反之, 若(2.3.1)$'$ 成立,则对于任意实数 x_1, y_1, x_2, y_2(其中 $x_1<x_2, y_1<y_2$)有

$$P\{x_1\leqslant\xi<x_2\}=F_1(x_2)-F_1(x_1);$$

$$P\{y_1\leqslant\eta<y_2\}=F_2(y_2)-F_2(y_1),$$

将上式左、右两端分别相乘,即得

$$P\{x_1\leqslant\xi<x_2\}P\{y_1\leqslant\eta<y_2\}=F_1(x_2)F_2(y_2)$$

① 设 ξ,η 的一切可能值分别为 $x_1,x_2,\cdots,x_i,\cdots$ 与 $y_1,y_2,\cdots,y_k,\cdots$.

$$-F_1(x_1)F_2(y_2)-F_1(x_2)F_2(y_1)+F_1(x_1)F_2(y_1)$$

$$=P\{x_1\leqslant\xi<x_2,y_1\leqslant\eta<y_2\}. \tag{2.3.3}$$

应用第一章性质1.3.4(关于概率连续性的定理)于(2.3.3)式，令$x_2\downarrow x_1, y_2\downarrow y_1$,即得

$$P\{\xi=x_1,\eta=y_1\}=P\{\xi=x_1\}P\{\eta=y_1\}.$$

对于密度函数$f(x,y)$为处处连续的连续型随机变数，条件(2.3.1)′等价于:

$$f(x,y)=f_1(x)\cdot f_2(y). \tag{2.3.4}$$

事实上，若(2.3.1)′式成立，则将(2.3.1)′式右端对x及y求导数,便可得到

$$\frac{\partial^2F(x,y)}{\partial x\partial y}=F'_1(x)F'_2(y)=f_1(x)f_2(y).$$

根据假设再将上式积分,并利用分布函数与密度函数的关系,便可得到

$$F(x,y)=F_1(x)F_2(y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f_1(u)f_2(v)dvdu$$

$$=\int_{-\infty}^{x}\int_{-\infty}^{y}\frac{\partial^2F(u,v)}{\partial u\partial v}dvdu.$$

从而由密度函数定义可知，$F(x,y)$有密度函数$f(x,y)\geqslant0$存在，且

$$f(x,y)=\frac{\partial^2F(x,y)}{\partial x\partial y}=f_1(x)f_2(y).$$

即(2.3.4)式成立.

反之,若(2.3.4)式成立,则有

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dxdy=\int_{-\infty}^{x}\int_{-\infty}^{y}f_1(x)f_2(y)dxdy$$

$$=\int_{-\infty}^{x}f_1(x)dx\cdot\int_{-\infty}^{y}f_2(y)dy=F_1(x)F_2(y).$$

即(2.3.1)′式成立.

若两个随机变数 ξ,η 不是独立的，我们就称它们是相依的。

在例 2.2.5 中，由于 ξ,η 的概率函数有

$$p(i,k)=\frac{1}{n^2},\qquad p(i,\cdot)=\frac{1}{n},$$

$$p(\cdot,k)=\frac{1}{n},$$

故对任意 $i,k=1,2,\cdots$，有

$$p(i,k)=p(i,\cdot)p(\cdot,k).$$

因此，ξ,η 是相互独立的.

但是在例 2.2.6 中，由于 $p(i,k)\neq p(i,\cdot)p(\cdot,k)$（如 $p(0,0)=\frac{1}{4}$，而 $P_\xi(0)=P(0,\cdot)=\frac{1}{2}$，$P_\eta(0)=p(\cdot,0)=\frac{13}{24}$，故 $p(0,0)=\frac{1}{4}$，$P_\xi(0)P_\eta(0))=\frac{13}{24}\cdot\frac{1}{2}=\frac{13}{48}$. 因此，$\xi,\eta$ 是相依的.

例 2.3.1 设二维随机变数 (ξ,η) 的联合密度函数为

$$f(x,y)=\begin{cases}e^{-(x+y)}, & 若\ x>0,y>0;\\ 0, & 其它.\end{cases}$$

求得：当 $x>0$ 时

$$f_\xi(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_0^{\infty}e^{-(x+y)}dy=e^{-x};$$

当 $x\leqslant 0$ 时

$$f_\xi(x)=0.$$

同样

$$f_\eta(y)=\begin{cases}e^{-y}, 若\ y>0;\\ 0, 若\ y\leqslant 0.\end{cases}$$

由此知

$$f(x,y)=f_\xi(x)f_\eta(y).$$

即 ξ,η 是相互独立的.

我们也可以从例 2.2.4 中看到，对于二维正态分布随机变数

(ξ,η)，由于 $f(x,y)\neq f_\xi(x)f_\eta(y)$，因此它们是相依的. 但当 $r=0$ 时，则可看到 $f(x,y)=f_\xi(x)f_\eta(y)$，因而，ξ,η 是独立的. 在下一章我们还会看到，逆命题也是成立的. 即当 (ξ,η) 的联合分布函数是服从例 2.2.4 的正态分布且相互独立时，则 $r=0$.

随机变数的独立性，可推广到有限个或可列无穷个的情形.

定义 2.3.2 设 $\xi_1,\xi_2,\cdots,\xi_n$ 是 n 个随机变数，$F(x_1,x_2,\cdots,x_n)$ 及 $F_i(x_i)$ 分别为 $(\xi_1,\xi_2,\cdots,\xi_n)$，$\xi_i(i=1,2,\cdots,n)$ 的分布函数，若对任意实数 $x_1,x_2,\cdots,x_n$，

$$F(x_1,x_2,\cdots,x_n)=F_1(x_1)F_2(x_2)\cdots F_n(x_n), \tag{2.3.5}$$

则称 $\xi_1,\xi_2,\cdots,\xi_n$ 是相互独立的.

定理 2.3.1 设 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立，则对任意 $k(2\leqslant k\leqslant n)$ 个随机变数 $\xi_{i_1},\xi_{i_2},\cdots\xi_{i_k}$ 也是相互独立的.

证 不失一般性，设 $i_1=1,i_2=2,\cdots,i_k=k$. 由相互独立性假设及(2.2.11)式，我们有

$$\begin{aligned}
&F_{1,2,\cdots,k}(x_1,x_2,\cdots,x_k)\\
&=F(x_1,x_2,\cdots,x_k,\overbrace{\infty,\infty,\cdots,\infty}^{n-k\text{个}})\\
&=F_1(x_1)F_2(x_2)\cdots F_k(x_k)F_{k+1}(\infty)\cdots F_n(\infty)\\
&=F_1(x_1)F_2(x_2)\cdots F_k(x_k)①. \quad ▌
\end{aligned}$$

定义 2.3.3 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 为一随机变数序列，若对任意 $k(k=2,\cdots,n,\cdots)$ 个随机变数 $\xi_{i_1},\xi_{i_2},\cdots,\xi_{i_k}$ 的联合分布函数

$$\begin{aligned}
&F_{i_1,i_2,\cdots,i_k}(x_{i_1},x_{i_2},\cdots,x_{i_k})\\
&=F_{i_1}(x_{i_1})F_{i_2}(x_{i_2})\cdots F_{i_k}(x_{i_k}).
\end{aligned} \tag{2.3.6}$$

其中 $F_{i_r}(x_{i_r}),r=1,2,\cdots,k$，为 ξ_{i_r} 的分布函数，则称随机变数序

---

① 这里 $F(x_1,x_2,\cdots,x_k,\infty,\infty,\cdots,\infty)=F_1(x_1)\cdots F_k(x_k)F_{k+1}(\infty)\cdots F_n(\infty)$ 是一种简写形式. 如果详细地写出来，则为 $\lim\limits_{x_{k+1},\cdots,x_n\to\infty} F(x_1,x_2,\cdots,x_k,x_{k+1},\cdots,x_n)=\lim\limits_{x_{k+1},\cdots,x_n\to\infty} F_1(x_1)\cdots F_k(x_k)F_{k+1}(x_{k+1})\cdots F_n(x_n)$.

列是相互独立的随机变数序列.

类似地，对于若干个随机向量，也能定义它们的相互独立性. 例如，设 η_1,η_2,η_3 分别为 n_1 维，n_2 维，n_3 维的随机向量；$F_{\eta_1}(y_1^{(1)},\cdots,y_{n_1}^{(1)})$，$F_{\eta_2}(y_1^{(2)},\cdots,y_{n_2}^{(2)})$，$F_{\eta_3}(y_1^{(3)},\cdots,y_{n_3}^{(3)})$ 分别为 η_1,η_2,η_3 的分布函数. 又记 (η_1,η_2,η_3) 的联合分布函数为 $F_{\eta_1,\eta_2,\eta_3}(y_1^{(1)},\cdots,y_{n_1}^{(1)};y_1^{(2)},\cdots,y_{n_2}^{(2)};y_1^{(3)},\cdots,y_{n_3}^{(3)})$；其中 $y_j^{(i)}$ 均为任意的实数，我们可以作出独立的定义如下：

若任取实数 $y_j^{(i)}$，都有 $F_{\eta_1,\eta_2,\eta_3}(y_1^{(1)},\cdots,y_{n_1}^{(1)};y_1^{(2)},\cdots,y_{n_2}^{(2)};y_1^{(3)},\cdots,y_{n_3}^{(3)})=F_{\eta_1}(y_1^{(1)},\cdots,y_{n_1}^{(1)})\cdot F_{\eta_2}(y_1^{(2)},\cdots,y_{n_2}^{(2)})\cdot F_{\eta_3}(y_1^{(3)},\cdots,y_{n_3}^{(3)})$ 成立，则称 η_1,η_2,η_3 是相互独立的.

对于随机变数是否相互独立，需要验证(2.3.1)′式或(2.3.5)式或(2.3.6)式. 至于独立性的直观意义，还不能从上述那些式子里得到较清晰的认识. 下面我们引进条件分布这一概念，进一步认识随机变数独立性这一重要的概念.

下列命题常常是有用的

定理 2.3.2　设有 $n_1+n_2+\cdots+n_k$ 个随机变数 $\xi_{11},\cdots,\xi_{1n_1},\xi_{21},\cdots,\xi_{2n_2},\cdots,\xi_{k1},\cdots\xi_{kn_k}$. 若 φ_i 是 n_i 个实变数的实值连续函数，令

$$\eta_i=\varphi_i(\xi_{i1},\cdots,\xi_{in_i})\quad(i=1,2,\cdots,k),$$

则

1° $\eta_1,\cdots,\eta_k$ 必是随机变数；

2° 若加设所有 ξ_{ij} 相互独立，则 $\eta_1,\cdots,\eta_k$ 也相互独立.

此定理的证明需用到较专门的知识(可参看[1] § 2.7)，这里从略.

二、条件分布

仿照条件概率定义，我们可以定义两个随机变数的条件分布. 下面着重讨论随机变数 ξ,η 为离散型或连续型时的情形.

(1) 设 ξ,η 为离散型随机变数，$p(i,k),p(i,\cdot),p(\cdot,k)$ 分别表示 (ξ,η)，ξ 及 η 的概率函数. 假设 $p(i,\cdot)>0,p(\cdot,k)>0$，我们称

$$\frac{p(i,k)}{p(i,\cdot)}\qquad(k=1,2,\cdots)\tag{2.3.7}$$

为 $\xi = x_i$ 条件下随机变数 η 的条件概率函数，并以 $P\{\eta = y_k/\xi = x_i\}$ 表示．如不产生混乱，简记为 $p(k/i)$．同样称

$$\frac{p(i,k)}{p(\cdot,k)} \qquad (i=1,2,\cdots) \tag{2.3.8}$$

为 $\eta = y_k$ 条件下随机变数 ξ 的条件概率函数，并以 $P\{\xi = x_i/\eta = y_k\}$ 表示，或简记为 $p(i/k)$.

如例2.2.5，那里

$$p(k/i) = \frac{p(i,k)}{p(i,\cdot)} = \frac{1/n^2}{1/n} = \frac{1}{n} \qquad (k=1,2,\cdots,n).$$

同理

$$p(i/k) = \frac{1}{n} \qquad (i=1,2,\cdots,n).$$

又如例 2.2.6，取 $i=0$，则

$$p(k/0) = \begin{cases} \dfrac{p(0,0)}{p(0,\cdot)} = \dfrac{1/4}{1/2} = \dfrac{1}{2} & \text{当 } k=0, \\[2ex] \dfrac{p(0,1)}{p(0,\cdot)} = \dfrac{1/4}{1/2} = \dfrac{1}{2} & \text{当 } k=1, \end{cases}$$

取 $i=1$，则

$$p(k/1) = \begin{cases} \dfrac{p(1,0)}{p(1,\cdot)} = \dfrac{1/6}{7/24} = \dfrac{4}{7} & \text{当 } k=0, \\[2ex] \dfrac{p(1,1)}{p(1,\cdot)} = \dfrac{1/8}{7/24} = \dfrac{3}{7} & \text{当 } k=1, \end{cases}$$

取 $i=2$，则

$$p(k/2) = \begin{cases} \dfrac{p(2,0)}{p(2,\cdot)} = \dfrac{1/8}{5/24} = \dfrac{3}{5} & \text{当 } k=0, \\[2ex] \dfrac{p(2,1)}{p(2,\cdot)} = \dfrac{1/12}{5/24} = \dfrac{2}{5} & \text{当 } k=1. \end{cases}$$

还可求得，取 $k=0$ 时，则

$$
p(i/0)=\begin{cases}\dfrac{p(0,0)}{p(\cdot,0)}=\dfrac{1/4}{13/24}=\dfrac{6}{13} & \text{当 } i=0, \\ \dfrac{p(1,0)}{p(\cdot,0)}=\dfrac{1/6}{13/24}=\dfrac{4}{13} & \text{当 } i=1, \\ \dfrac{p(2,0)}{p(\cdot,0)}=\dfrac{1/8}{13/24}=\dfrac{3}{13} & \text{当 } i=2,\end{cases}
$$

取 $k=1$ 时，则

$$
p(i/1)=\begin{cases}\dfrac{p(0,1)}{p(\cdot,1)}=\dfrac{1/4}{11/24}=\dfrac{6}{11} & \text{当 } i=0, \\ \dfrac{p(1,1)}{p(\cdot,1)}=\dfrac{1/8}{11/24}=\dfrac{3}{11} & \text{当 } i=1, \\ \dfrac{p(2,1)}{p(\cdot,1)}=\dfrac{1/12}{11/24}=\dfrac{2}{11} & \text{当 } i=2.\end{cases}
$$

我们称

$$
\frac{\sum\limits_{x_i<x} p(i,k)}{p(\cdot,k)} \tag{2.3.9}
$$

为 $\eta=y_k$ 条件下 ξ 的条件分布函数，并且用 $P\{\xi<x/\eta=y_k\}$ 表示，或简写为 $F(x/y_k)$.

同样称

$$
\frac{\sum\limits_{y_k<y} p(i,k)}{p(i,\cdot)} \tag{2.3.10}
$$

为 $\xi=x_i$ 条件下，η 的条件分布函数，并且用 $P\{\eta<y/\xi=x_i\}$ 表示，或简写为 $F(y/x_i)$.

从上面我们看到，$F(x/y_k)$ 及 $F(y/x_i)$ 不仅是 x 或 y 的函数，且依赖于 y_k 或 x_i. 即与随机变数 η 或 ξ 取值有关.

(2) 若 ξ,η 为连续型随机变数，$f(x,y)$，$f_1(x)$ 及 $f_2(y)$ 分别为 (ξ,η)，ξ 及 η 的密度函数，且对任意 x 及 y

$$
f_1(x)=\int_{-\infty}^{\infty} f(x,v)\,dv>0;
$$

$$f_2(y)=\int_{-\infty}^{\infty} f(u,y)\,du>0.$$

则称

$$\frac{\int_{-\infty}^{x} f(u,y)\,du}{\int_{-\infty}^{\infty} f(u,y)\,du}=\frac{\int_{-\infty}^{x} f(u,y)\,du}{f_2(y)} \tag{2.3.11}$$

为在条件 $\eta=y$ 下，ξ 的条件分布函数，以 $P\{\xi<x/\eta=y\}$ 表示，或简记为 $F(x/y)$.

称

$$\frac{\int_{-\infty}^{y} f(x,v)\,dv}{\int_{-\infty}^{\infty} f(x,v)\,dv}=\frac{\int_{-\infty}^{y} f(x,v)\,dv}{f_1(x)} \tag{2.3.12}$$

为在条件 $\xi=x$ 下，η 的条件分布函数，以 $P\{\eta<y/\xi=x\}$ 表示，或简记为 $F(y/x)$.

令 $$f(y/x)=\frac{f(x,y)}{f_1(x)},\quad f(x/y)=\frac{f(x,y)}{f_2(y)}, \tag{2.3.13}$$

则(2.3.11)及(2.3.12)两式可改写为：

$$\left.\begin{aligned}F(x/y)&=\int_{-\infty}^{x} f(u/y)\,du,\\ F(y/x)&=\int_{-\infty}^{y} f(v/x)\,dv.\end{aligned}\right\} \tag{2.3.14}$$

我们称 $f(x/y)$ 为在条件 $\eta=y$ 下，ξ 的条件密度函数，$f(y/x)$ 为在条件 $\xi=x$ 下，η 的条件密度函数.

例 2.3.2 设 (ξ,η) 为服从二维正态分布 $N(0,1;0,1;r)$ 的随机变数，试求 $f(y/x)$ 及 $f(x/y)$.

解 因 $f(x,y)=\frac{1}{2\pi\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)\right\}$,

根据(2.3.13)式和(2.2.4)及(2.2.5)两式

$$f(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{1}{2\pi\sqrt{1-r^2}}$$

$$\cdot\exp\left\{\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)\right\}\Big/\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}e^{-\frac{1}{2}\left(\frac{y-rx}{\sqrt{1-r^2}}\right)^2}.$$

同理可得

$$f(x/y)=\frac{f(x,y)}{f_2(y)}=\frac{1}{2\pi\sqrt{1-r^2}}$$

$$\cdot\exp\left\{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)\right\}\Big/\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}e^{-\frac{1}{2}\left(\frac{x-ry}{\sqrt{1-r^2}}\right)^2}.$$

上二式说明：在条件 $\xi=x$ 下，η 的条件密度函数为正态分布 $N(rx,\sqrt{1-r^2})$. 或在条件 $\eta=y$ 下，ξ 的条件密度函数为正态分布 $N(ry,\sqrt{1-r^2})$. 后者的几何意义为：$f(x/y)=\dfrac{f(x,y)}{f_2(y)}$ 图形是由联合密度函数所描述的曲面被垂直于 y 轴且平行于 XOZ 平面的过 $(0,y,0)$ 点的平面所切割，然后乘以常数因子 $1/f_2(y)$ 得到的曲线. 当 y 变动时，得到一族形状类似于图 2.3.1 的曲线. 可以证明，由(2.3.9)式～(2.3.12)式所定义的条件分布函数，也满足定理 2.1.2 关于分布函数的三个性质.

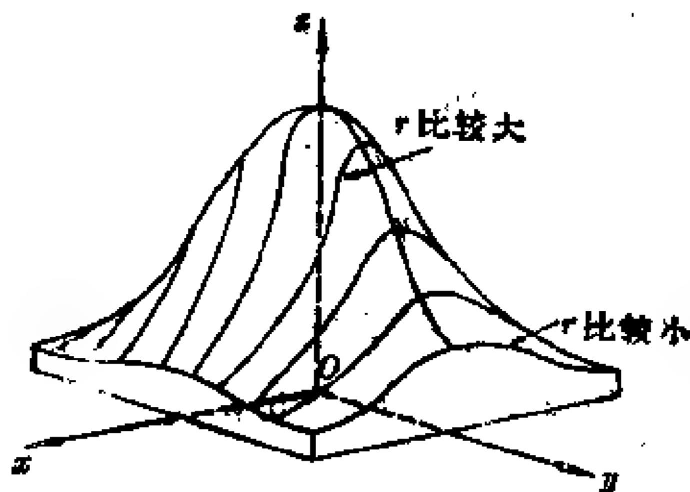


图 2.3.1

*(3) 一般情形

设(ξ,η)的联合分布函数为$F(x,y)$，记

$$P(\eta<y\mid\xi=x)=\lim_{\alpha,\beta\to0}\frac{F(x+\beta,y)-F(x-\alpha,y)}{F(x+\beta,\infty)-F(x-\alpha,\infty)}$$

$$(\alpha>0,\beta>0).\qquad(2.3.15)$$

若右边极限存在，则称之为在条件$\xi=x$下，η的条件分布函数.

注意，在适当条件下，(2.3.15)式确是(2.3.12)式之推广. 亦即，当$F(x,y)$有联合密度函数$f(x,y)\geqslant0$，使

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\,du\,dv$$

时，则在适当条件下，由(2.3.15)式可得出(2.3.12)式.

事实上，此时，由(2.3.15)式得

$$P(\eta<y\mid\xi=x)=\lim_{\alpha,\beta\to0}\frac{\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{y}f(u,v)\,du\,dv}{\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{\infty}f(w,z)\,dw\,dz}$$

$$=\lim_{\alpha,\beta\to0}\frac{\frac{1}{\alpha+\beta}\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{y}f(u,v)\,du\,dv}{\frac{1}{\alpha+\beta}\int_{x-\alpha}^{x+\beta}\int_{-\infty}^{\infty}f(w,z)\,dw\,dz}=\frac{\int_{-\infty}^{y}f(x,v)\,dv}{\int_{-\infty}^{\infty}f(x,v)\,dv}$$

$$=\frac{\int_{-\infty}^{y}f(x,v)\,dv}{f_1(x)}.$$

(要上式等号成立，只须$\int_{-\infty}^{y}f(x,v)\,dv$及$\int_{-\infty}^{\infty}f(x,v)\,dv$在$x$点连续即可.)

最后，让我们指出，若ξ,η相互独立，则η在$\xi=x$的条件下的条件分布函数与η的(无条件)分布函数$F_2(y)$一样，这正是我们所期望的结果. 它告诉我们关于随机变数独立性的直观解释.

事实上，若ξ,η独立，由(2.3.15)式，则

$$P(\eta<y|\xi=x)=\lim_{\alpha,\beta\to 0}\frac{F(x+\beta,y)-F(x-\alpha,y)}{F(x+\beta,\infty)-F(x-\alpha,\infty)}$$

$$=\lim_{\alpha,\beta\to 0}\frac{F_1(x+\beta)F_2(y)-F_1(x-\alpha)F_2(y)}{F_1(x+\beta)-F_1(x-\alpha)}=F_2(y).$$

(对于 $P\{\xi<x|\eta=y\}$ 的情形，可同样讨论.)

§2.4 随机变数的函数及其分布函数

在概率论与数理统计的许多问题中，常需要考虑随机变数的函数或其变换，特别在数理统计中更为重要.

例 2.4.1 由统计物理知道，气体分子运动速度的绝对值 ξ 服从马克斯威尔分布，其密度函数为

$$f_\xi(x)=\begin{cases}\dfrac{4x^2}{\alpha^3\sqrt{\pi}}e^{-\frac{x^2}{\alpha^2}}, & \text{当 } x>0;\\ 0, & \text{当 } x\leqslant 0,\end{cases}\qquad(\alpha>0),$$

试求分子运动动能 $\eta=\frac{1}{2}m\xi^2$ 所服从的分布(其中 m 表分子的质量).

解 η 的分布函数是

$$F_\eta(x)=P\left(\frac{1}{2}m\xi^2<x\right)=P\left(0\leqslant\xi<\sqrt{\frac{2x}{m}}\right)$$

$$=\int_0^{\sqrt{\frac{2x}{m}}}\frac{4t^2}{\alpha^3\sqrt{\pi}}e^{-\frac{t^2}{\alpha^2}}dt,\quad \text{当 } x>0 \text{ 时.}$$

故它的密度函数是

$$f_\eta(x)=F'_\eta(x)=\frac{4\left(\dfrac{2x}{m}\right)}{\alpha^3\sqrt{\pi}}e^{-\frac{2x}{m\alpha^2}}\cdot\frac{1}{\sqrt{2mx}}$$

$$=\frac{4\sqrt{2x}}{\left(\alpha m^{\frac{1}{2}}\right)^3\sqrt{\pi}}e^{-\frac{2x}{m\alpha^2}},\quad \text{当 } x>0 \text{ 时.}$$

显然，当 $x\leqslant 0$ 时，$F_\eta(x)=0$. 故可以有：

$$f_\eta(x)=0, \text{当 } x\leqslant 0 \text{ 时.}$$

现在来讨论求一般的随机变数函数的分布函数问题. 这样一类问题,一般可叙述如下:

设$(\xi_1,\xi_2,\cdots,\xi_n)$为$n$维随机变数, 若已知其联合分布. 又设有$k$个$\xi_1,\xi_2,\cdots,\xi_n$的函数:

$$\eta_1=g_1(\xi_1,\xi_2,\cdots,\xi_n),$$
$$\cdots\cdots\cdots\cdots$$
$$\eta_k=g_k(\xi_1,\xi_2,\cdots,\xi_n),$$

其中$g_i(\xi_1,\cdots,\xi_n)$ $(i=1,2,\cdots,k)$均为n元连续函数, 则由定理2.3.2知,$\eta_1,\cdots,\eta_k$均为随机变数. 现问$(\eta_1,\eta_2,\cdots,\eta_k)$的联合分布如何求出?

这问题的一般解答如下:

$$\begin{aligned}&F_{\eta_1,\eta_2,\cdots,\eta_k}(y_1,y_2,\cdots,y_k)=P\{\eta_1<y_1,\eta_2<y_2,\cdots,\eta_k<y_k\}\\&=P\{g_1(\xi_1,\xi_2,\cdots,\xi_n)<y_1,\cdots,g_k(\xi_1,\xi_2,\cdots,\xi_n)<y_k\}\\&=P\{(\xi_1,\xi_2,\cdots,\xi_n)\in\bigcap_{i=1}^{k}g_i^{-1}[(-\infty,y_i)]\}.\end{aligned}\tag{2.4.1}$$

这里事件$\{(\xi_1,\xi_2,\cdots,\xi_n)\in\bigcap_{i=1}^{k}g_i^{-1}[(-\infty,y_i)]\}$是事件$\{g_1(\xi_1,\xi_2,\cdots,\xi_n)<y_1,\cdots,g_k(\xi_1,\xi_2,\cdots,\xi_n)<y_k\}$的另一种表达方式. 进一步, 当$(\xi_1,\xi_2,\cdots,\xi_n)$为连续型的时候, 即具有联合密度函数$f_\xi(x_1,x_2,\cdots,x_n)$时,

$$F_{\eta_1,\eta_2,\cdots,\eta_k}(y_1,y_2,\cdots,y_k)=\int_D\cdots\int f_\xi(x_1,x_2,\cdots,x_n)\,dx_1\cdots dx_n.\tag{2.4.2}$$

其中D由不等式$g_i(x_1,\cdots,x_n)<y_i$, $(i=1,2,\cdots,k)$决定.

当$(\xi_1,\xi_2,\cdots,\xi_n)$为离散型, 则上述分布函数为一个在$D$的范围上求的$n$重和式.

表面看来,(2.4.1)式(或(2.4.2)式,或离散型的n重和式)已解决了全部问题. 但是进一步考虑着手于实际计算时, 要真的把

$F_{\eta_1,\eta_2,\cdots,\eta_k}(y_1,y_2,\cdots,y_k)$的解析表达式求出来，却往往并不容易．原因在于$D$之形式往往很复杂，因此引起计算上的困难．下面仅讨论几种重要特殊情形．

一、和的分布

设(ξ_1,ξ_2)的联合分布函数为$F_{\xi_1,\xi_2}(x_1,x_2)$，则$\xi_1+\xi_2$的分布函数为

$$F_{\xi_1+\xi_2}(x)=P(\xi_1+\xi_2<x).$$

当$F_{\xi_1,\xi_2}(x_1,x_2)$有密度函数$f_{\xi_1,\xi_2}(x_1,x_2)$时，由(2.4.2)式得

$$F_{\xi_1+\xi_2}(x)=\iint\limits_{x_1+x_2<x} f_{\xi_1,\xi_2}(x_1,x_2)\,dx_1dx_2.$$

积分区域如图2.4.1的阴影部分．

令

$$\begin{cases}z=x_1+x_2,\\ x_1=x_1,\end{cases}$$

有

$$\frac{\partial(x_1,x_2)}{\partial(x_1,z)}=\begin{vmatrix}1 & 0\\ -1 & 1\end{vmatrix}=1.$$

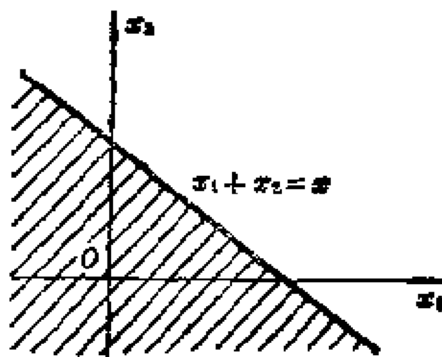


图 2.4.1

故得

$$F_{\xi_1+\xi_2}(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_1,z-x_1)\,dx_1dz.$$

因此，$\xi_1+\xi_2$之分布密度函数为

$$f_{\xi_1+\xi_2}(x)=\int_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_1,x-x_1)\,dx_1. \tag{2.4.3}$$

特别，当ξ_1,ξ_2相互独立时，有

$$f_{\xi_1+\xi_2}(x)=\int_{-\infty}^{\infty} f_{\xi_1}(x_1)\cdot f_{\xi_2}(x-x_1)\,dx_1. \tag{2.4.4}$$

(2.4.4)式右边之积分我们常常称为函数$f_{\xi_1}(x)$与$f_{\xi_2}(x)$的卷积．记作

$$f_{\xi_1}(x)*f_{\xi_2}(x)\equiv\int_{-\infty}^{\infty}f_{\xi_1}(x_1)\cdot f_{\xi_2}(x-x_1)\,dx_1 \qquad (2.4.5)$$

由此得一重要结论：

定理 2.4.1 若 ξ_1，ξ_2 相互独立且均为连续型随机变数，则 $\xi_1+\xi_2$ 也是连续型随机变数，并且其密度函数为它们二个密度函数之卷积.

对于离散型随机变数之相加，也可以有类似的卷积求概率公式. 例如，设 ξ，η 为离散型随机变数，分别有概率分布列

$$P(\xi=a_i)=p_i,\qquad \sum_{i=1}^{\infty}p_i=1,$$

$$P(\eta=b_k)=p_k^*,\qquad \sum_{k=1}^{\infty}p_k^*=1.$$

则 $\xi+\eta$ 的分布函数是

$$F_{\xi+\eta}(x)=P(\xi+\eta<x)=\sum_{a_i+b_k<x}p_{ik}$$

$$=\sum_{k=1}^{\infty}\sum_{a_i<x-b_k}p_{ik};$$

其中记 $p_{ik}=P(\xi=a_i,\eta=b_k)$，上面的第一和式是 k 固定时，对所有满足不等式 $a_i<x-b_k$ 的 i 来求和. 当 ξ 与 η 独立时，上式变为

$$F_{\xi+\eta}(x)=\sum_{k=1}^{\infty}\sum_{a_i<x-b_k}p_i\cdot p_k^*.$$

值得注意的是，上面式子还可以有别的表示式，由于 $\xi+\eta$ 至多只取可数个值 $\{a_i+b_k\}_{i,k=1}^{\infty}$（其中有些值可能相同），因此，

$$F_{\xi+\eta}(x)=\sum_{t<x}\sum_{i=1}^{\infty}P(\xi=a_i,\eta=t-a_i),$$

其中第二个和式至多是可列个相加，因为

$$P(\xi=a_i,\eta=t-a_i)\neq0$$

的 t 至多可列个而已. 进一步，若将 $\{a_i+b_k\}_{i,k=1}^{\infty}$ 全部排列成

$\{c_i\}_{i=1}^{\infty}$，则还可以有表示式

$$P(\xi+\eta=c_i)=\sum_{i=1}^{\infty}P(\xi=a_i,\eta=c_i-a_i).$$

这个公式类似于连续型的公式(2.4.4)．特别，当 ξ 与 η 相互独立时，有

$$P(\xi+\eta=c_i)=\sum_{i=1}^{\infty}P_i\cdot P(\eta=c_i-a_i).$$

这个公式类似于连续型的公式(2.4.5)．也称它为 ξ 与 η 的两个分布列的卷积．

下面利用定理 2.4.1 去证明：k 个相互独立的服从Γ-分布的随机变数之和，仍服从Γ-分布．

定理 2.4.2 若 $\xi_1,\xi_2,\cdots,\xi_k$ 相互独立，且 ξ_i 服从Γ-分布 $\Gamma(\alpha_i,\beta)$ $(i=1,2,\cdots,k)$，则它们的和 $\xi_1+\xi_2+\cdots+\xi_k$ 服从Γ-分布$\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_k+k-1,\beta)$．

证 先证 $k=2$ 的情形．由设 ξ_1,ξ_2之密度函数分别是：

$$f_1(x_1)=\frac{1}{\beta^{\alpha_1+1}\,\Gamma(\alpha_1+1)}x_1^{\alpha_1}e^{-x_1/\beta},\text{当 } x_1\in(0,\infty)\text{时},$$

$$f_2(x_2)=\frac{1}{\beta^{\alpha_2+1}\,\Gamma(\alpha_2+1)}x_2^{\alpha_2}e^{-x_2/\beta},\text{当 } x_2\in(0,\infty)\text{时}.$$

由定理 2.4.1，得

$$\begin{aligned}
f_{\xi_1+\xi_2}(x)&=\int_{-\infty}^{\infty}f_1(x_1)\cdot f_2(x-x_1)\,dx_1\\
&=\frac{1}{\beta^{\alpha_1+\alpha_2+2}\Gamma(\alpha_1+1)\Gamma(\alpha_2+1)}\int_0^x x_1^{\alpha_1}(x-x_1)^{\alpha_2}e^{-\frac{1}{\beta}(x_1+x-x_1)}\,dx_1\\
&=\frac{1}{\beta^{(\alpha_1+\alpha_2+1)+1}\Gamma(\alpha_1+1)\Gamma(\alpha_2+1)}e^{-\frac{x}{\beta}}\int_0^x x_1^{\alpha_1}(x-x_1)^{\alpha_2}dx_1\\
&=\frac{1}{\beta^{(\alpha_1+\alpha_2+1)+1}\cdot\Gamma(\alpha_1+\alpha_2+1+1)}e^{-\frac{x}{\beta}}\cdot x^{\alpha_1+\alpha_2+1}.
\end{aligned}$$

即得 $\xi_1+\xi_2$服从$\Gamma(\alpha_1+\alpha_2+1,\beta)$．最后一等式，是由于公式

$$B(m,n)=\int_0^1 x_1^{m-1}(1-x_1)^{n-1}dx_1=\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$$

及

$$\int_0^x x_1^{\alpha_1}(x-x_1)^{\alpha_2}dx_1=\int_0^x x_1^{\alpha_1}x^{\alpha_2}\left(1-\frac{x_1}{x}\right)^{\alpha_2}dx_1$$

$$=\int_0^x\left(\frac{x_1}{x}\right)^{\alpha_1}x^{\alpha_1+\alpha_2+1}\left(1-\frac{x_1}{x}\right)^{\alpha_2}d\left(\frac{x_1}{x}\right)$$

$$=\left(\int_0^1 y^{\alpha_1}(1-y)^{\alpha_2}dy\right)\cdot x^{\alpha_1+\alpha_2+1}.$$

由归纳法，易得对任意正整数 k，定理亦成立. ▌

后面我们将要看到：在推导于统计中有重要应用的χ^2-分布，t-分布，F-分布时，运用定理 2.4.2 将带来极大的方便.

再举一例说明独立随机变数和的密度函数公式 (2.4.4) 式之应用.

例 2.4.2 设 ξ_1,ξ_2 相互独立，且皆取值于[0,1]区间，服从均匀分布，试求 $\xi_1+\xi_2$ 的密度函数。

解 由设

$$f_{\xi_1}(x)=f_{\xi_2}(x)=\begin{cases}1, \text{当 } x\in[0,1],\\ 0, \text{其它}.\end{cases}$$

故

$$f_{\xi_1+\xi_2}(x)=\int_{-\infty}^{+\infty}f_{\xi_1}(x_1)\cdot f_{\xi_2}(x-x_1)dx_1=\int_0^1 f_{\xi_2}(x-x_1)dx_1$$

$$=\int_{x-1}^{x}f_{\xi_2}(y)dy.$$

如图 2.4.2，当 $0<x-1<1$ 时，上式等于：

$$\int_{x-1}^{1}1dy=2-x.$$

当 $0\leqslant x\leqslant 1$ 时，上式等于：

$$\int_0^x 1dy=x.$$

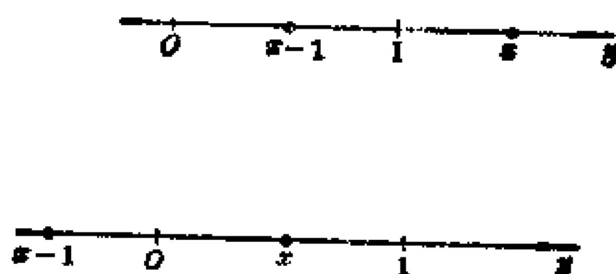


图 2.4.2

对于其它的 x，上式等于 0。

故

$$f_{\xi_1+\xi_2}(x)=\begin{cases}x, & 当\ 0\leqslant x\leqslant 1,\\ 2-x, & 当\ 1\leqslant x\leqslant 2,\\ 0, & 其它.\end{cases}$$

令 $z=f_{\xi_1+\xi_2}(x)$，它的图形如图 2.4.3，称它为三角形密度函数。

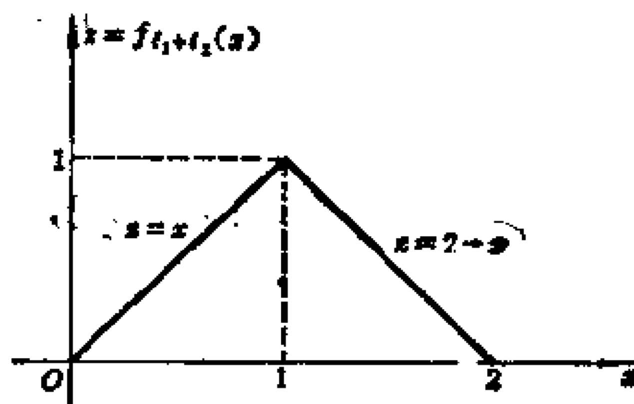


图 2.4.3

二、商的分布

设(ξ_1,ξ_2)的联合密度函数为 $f_{\xi_1,\xi_2}(x_1,x_2)$，则$\frac{\xi_1}{\xi_2}$（设 $\xi_2\neq 0$）的分布函数为

$$F_{\frac{\xi_1}{\xi_2}}(x)=P\left(\frac{\xi_1}{\xi_2}<x\right).$$

由(2.4.2)式，得

$$F_{\frac{\xi_1}{\xi_2}}(x)=\iint\limits_{\frac{x_1}{x_2}<x} f_{\xi_1,\xi_2}(x_1,x_2)\,dx_1dx_2.$$

积分区域如图 2.4.4 的阴影部分。

令

$$\begin{cases} z=\dfrac{x_1}{x_2}, \\ x_2=x_2. \end{cases}$$

有

$$\frac{\partial(x_1,x_2)}{\partial(z,x_2)}=\begin{vmatrix} x_2 & z \\ 0 & 1 \end{vmatrix}$$

$$=x_2.$$

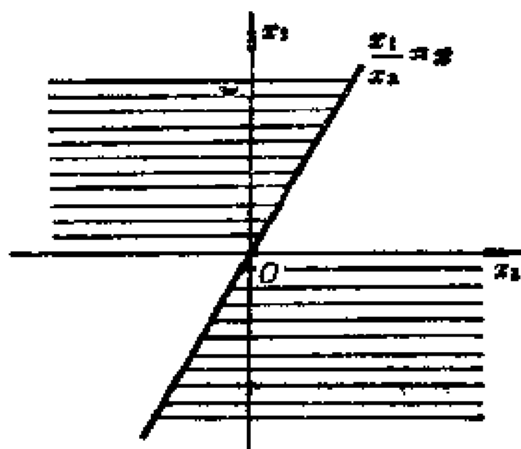


图 2.4.4

故

$$F_{\frac{\xi_1}{\xi_2}}(x)=\int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_2z,x_2)\,|x_2|\,dx_2dz.$$

亦即$\frac{\xi_1}{\xi_2}$的密度函数为

$$f_{\frac{\xi_1}{\xi_2}}(x)=\int_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_2x,x_2)\,|x_2|\,dx_2. \tag{2.4.6}$$

特别，当 ξ_1,ξ_2 独立时，有

$$f_{\frac{\xi_1}{\xi_2}}(x)=\int_{-\infty}^{\infty} f_{\xi_1}(x_2x)\cdot f_{\xi_2}(x_2)\,|x_2|\,dx_2. \tag{2.4.7}$$

例2.4.3 设 ξ，η 相互独立，其密度函数分别如下(如图 2.4.5)：

$$f_\xi(x)=\begin{cases} \dfrac{1}{2}, & x\in[1,3], \\ 0, & \text{其它}; \end{cases} \qquad f_\eta(y)=\begin{cases} e^{-(y-2)}, & y\in[2,\infty), \\ 0, & \text{其它}. \end{cases}$$

试求 $f_{\frac{\xi}{\eta}}(x)$.

解 $f_{\frac{\xi}{\eta}}(x)=\int_{-\infty}^{\infty} f_\xi(x_2x)\cdot f_\eta(x_2)\,|x_2|\,dx_2=\int_{\frac{1}{x}}^{\frac{3}{x}}\frac{1}{2}\,|x_2|\,f_\eta(x_2)\,dx_2.$

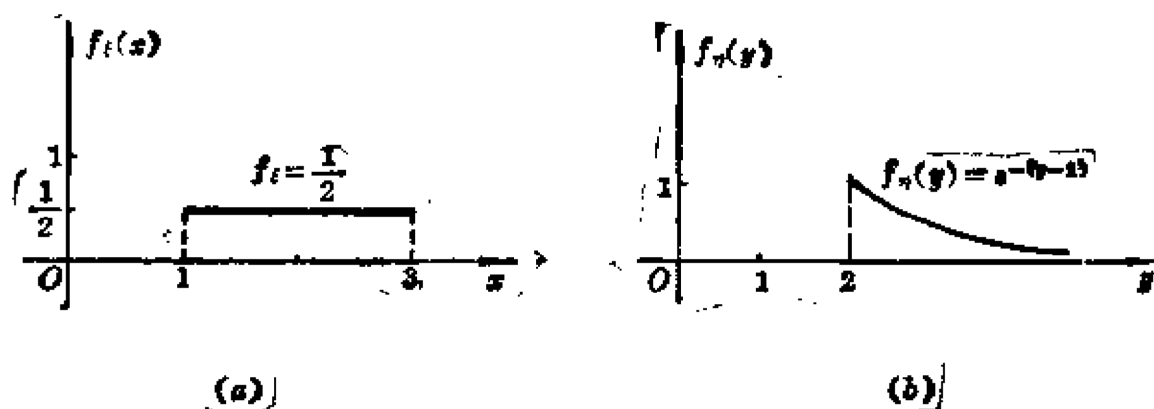


图 2.4.5

如图 2.4.6，当 $0<x\leqslant\frac{1}{2}$ 时，

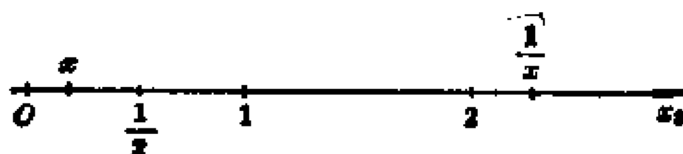


图 2.4.6

$$f_{\frac{\xi}{\eta}}(x)=\int_{\frac{1}{x}}^{\frac{3}{x}}\frac{1}{2}x_2e^{-(x_2-2)}dx_2$$

$$=-\frac{1}{2}e^{-(x_2-2)}\cdot x_2\Big|_{x_2=\frac{1}{x}}^{x_2=\frac{3}{x}}+\frac{1}{2}\int_{\frac{1}{x}}^{\frac{3}{x}}e^{-(x_2-2)}dx_2$$

$$=\frac{1}{2}\left[\frac{1}{x}e^{-(\frac{1}{x}-2)}-\frac{3}{x}e^{-(\frac{3}{x}-2)}\right]$$

$$-\frac{1}{2}e^{-(\frac{3}{x}-2)}+\frac{1}{2}e^{-(\frac{1}{x}-2)}$$

$$=\frac{1}{2x}e^2\left[e^{-\frac{1}{x}}(1+x)-e^{-\frac{3}{x}}(3+x)\right].$$

如图 2.4.7，

当 $\frac{1}{2}<x\leqslant\frac{3}{2}$ 时，

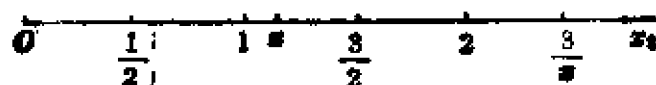


图 2.4.7

$$
\begin{aligned}
f_{\frac{\xi}{\eta}}(x) &= \int_2^{\frac{3}{x}} \frac{1}{2} x_2 e^{-(x_2-2)}\, dx_2 \\
&= -\frac{1}{2} e^{-(x_2-2)} x_2 \Big|_{x_2=2}^{x_2=\frac{3}{x}} - \frac{1}{2} e^{-(x_2-2)} \Big|_{x_2=2}^{x_2=\frac{3}{x}} \\
&= \frac{3}{2} - \frac{e^2}{2x} e^{-\frac{3}{x}}(3+x).
\end{aligned}
$$

最后得

$$
f_{\frac{\xi}{\eta}}(x) = \begin{cases} \dfrac{1}{2x} e^2 \left[e^{-\frac{1}{x}}(1+x) - e^{-\frac{3}{x}}(x+3) \right], & \text{当 } 0 < x \leqslant \dfrac{1}{2} \text{ 时;} \\ \dfrac{3}{2} - \dfrac{e^2}{2x} e^{-\frac{3}{x}}(3+x), & \text{当 } \dfrac{1}{2} < x \leqslant \dfrac{3}{2} \text{ 时;} \\ 0, & \text{其它.} \end{cases}
$$

用上面类似的方法，还可得到随机变数的差的分布与积的分布.

三、随机变数的线性变换与平方变换

设 ξ 为任一随机变数，$F_\xi(x)$ 为其分布函数，则

$$
\left.\begin{aligned}
&1^\circ \ \text{当 } a>0,\ -\infty<b<\infty \text{ 时，有} \\
&\quad F_{a\xi+b}(x) = F_\xi\left(\frac{x-b}{a}\right) \\
&\quad \text{当 } a<0,\ -\infty<b<\infty \text{ 时，有} \\
&\quad F_{a\xi+b}(x) = 1 - F_\xi\left(\frac{x-b}{a}+0\right)
\end{aligned}\right\} \qquad (2.4.8)
$$

(这里 $F(x+0) \equiv \lim\limits_{\substack{h\to 0 \\ h>0}} F(x+h)$).

特别地，当 $F_\xi(x)$ 有密度函数 $f_\xi(x)$ 时(此时 $F_\xi(x)$ 为连续)，有

$$f_{a\xi+b}(x)=\frac{1}{|a|}f_\xi\left(\frac{x-b}{a}\right),\quad a\neq 0. \tag{2.4.9}$$

当 ξ 为离散随机变数时，有

$$P_{a\xi+b}(x_i)=P_\xi\left(\frac{x_i-b}{a}\right).$$

$$2^\circ\quad F_{\xi^2}(x)=\begin{cases}F_\xi(\sqrt{x})-F_\xi(-\sqrt{x}+0), & \text{当 } x>0 \text{ 时;}\\ 0, & \text{其它.}\end{cases} \tag{2.4.10}$$

当 $F_\xi(x)$ 有连续的密度函数 $f_\xi(x)$（或 $f_\xi(x)$ 只有有限个间断点），则

$$f_{\xi^2}(x)=\frac{1}{2\sqrt{x}}[f_\xi(\sqrt{x})+f_\xi(-\sqrt{x})],\ (x>0). \tag{2.4.11}$$

证 先证 1°，$a>0$ 时.

$$F_{a\xi+b}(x)=P(a\xi+b<x)=P\left(\xi<\frac{x-b}{a}\right)=F_\xi\left(\frac{x-b}{a}\right).$$

若有密度函数，则

$$F_{a\xi+b}(x)=F_\xi\left(\frac{x-b}{a}\right)=\int_{-\infty}^{\frac{x-b}{a}}f_\xi(u)du.$$

令 $v=au+b$，即得

$$F_{a\xi+b}(x)=\frac{1}{a}\int_{-\infty}^{x}f_\xi\left(\frac{v-b}{a}\right)dv.$$

故

$$f_{a\xi+b}(x)=\frac{1}{a}f_\xi\left(\frac{x-b}{a}\right)\qquad (a>0).$$

同理可证 $a<0$ 时情形.

当 ξ 为离散随机变数时，则

$$P_{a\xi+b}(x_i)=P\{a\xi+b=x_i\}=P\left\{\xi=\frac{x_i-b}{a}\right\}=P_\xi\left(\frac{x_i-b}{a}\right).$$

其次证 2° 设 $x>0$

$$F_{\xi^2}(x)=P(\xi^2<x)=P(-\sqrt{x}<\xi<\sqrt{x})$$
$$=P(\xi<\sqrt{x})-P(\xi\leqslant-\sqrt{x})$$
$$=F_\xi(\sqrt{x})-F_\xi(-\sqrt{x}+0).$$
$$F_{\xi^2}(x)=0,\text{当 } x\leqslant 0.$$

当 $F_\xi(x)$ 有连续的密度函数 $f_\xi(x)$ 时，(或 $f_\xi(x)$ 有有限个间断点)则在 x 为 $f_\xi(x)$ 的连续点时

$$f_{\xi^2}(x)=\frac{d}{dx}F_{\xi^2}(x)=\frac{f_\xi(\sqrt{x})+f_\xi(-\sqrt{x})}{2\sqrt{x}},\qquad (x>0).$$ ■

例 2.4.4 1) 设 ξ 服从正态 $N(0,1)$ 分布，求 $a\xi+b$ 之密度函数.

2) 设 η 服从正态 $N(m,\sigma)$，求 $\xi=\frac{\eta-m}{\sigma}$ 的密度函数.

解 1) $f_{a\xi+b}(x)=\frac{1}{|a|}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-b)^2}{2a^2}}$，故 $a\xi+b$ 服从正态 $N(b,|a|)$.

2) $f_{\frac{\eta-m}{\sigma}}(x)=\sigma\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\sigma x+m-m)^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，故 $\xi=\frac{\eta-m}{\sigma}$ 服从正态 $N(0,1)$ (正态 $N(0,1)$ 有时称为标准正态 $N(0,1)$，而线性变换 $\frac{\eta-m}{\sigma}$ 称为对随机变数 η 的标准化，ξ 称为标准化随机变数). 此例说明服从标准正态 $N(0,1)$ 的随机变数作了线性变换后仍服从正态分布. 反过来，服从任一正态分布的随机变数也必定可以标准化，使其服从标准正态 $N(0,1)$.

例 2.4.5 设 ξ 服从正态 $N(0,1)$，求 ξ^2 之密度函数.

解 $$f_{\xi^2}(x)=\frac{1}{2\sqrt{x}}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}+\frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}\right)$$
$$=\frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{x}{2}},\qquad (x>0).\tag{2.4.12}$$

即 ξ^2 服从 Γ- 分布：$\Gamma\left(-\frac{1}{2},2\right)$.

四、χ^2-分布，t-分布，F-分布

应用上面讲过的知识，现在很容易推导出在统计中有着重要应用的三种分布.

定理 2.4.3（χ^2-分布） 若 n 个相互独立的随机变数 $\xi_1, \cdots, \xi_n$，均服从正态 $N(0,1)$，则 $\chi^2=\sum_{k=1}^{n}\xi_k^2$ 的密度函数为：

$$f_{\chi^2}(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & \text{当 } x\geqslant 0, \\ 0, & \text{其它,}\end{cases} \tag{2.4.13}$$

称(2.4.13)为自由度为 n 的 χ^2-密度函数(附录有表可查)①

证 由例2.4.5知，ξ 服从 $N(0, 1)$，则 ξ^2 服从 Γ- 分布 $\Gamma\left(-\frac{1}{2},2\right)$. 由定理 2.4.2 知，$\chi^2$ 服从 Γ-分布 $\Gamma\left(-\frac{n}{2}+n-1,2\right)$，即服从 $\Gamma\left(\frac{n}{2}-1,2\right)$，亦即(2.4.13)式成立. ▌

推论 1

$$f_{\frac{\chi^2}{n}}(x)=\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{nx}{2}} \quad (x\geqslant 0), \tag{2.4.15}$$

---

① 由例2.4.4知，若 η_i 服从正态 $N(m_i,\sigma_i)$，则 $\xi_i=\frac{\eta_i-m_i}{\sigma_i}$ 服从 $N(0,1)$. 故此定理亦可推广为：

若 n 个相互独立之随机变数 $\eta_1,\cdots,\eta_n$，分别服从 $N(m_i,\sigma_i)$，则

$$\chi^2=\sum_{i=1}^{n}\left(\frac{\eta_i-m_i}{\sigma_i}\right)^2 \tag{2.4.14}$$

的密度函数为(2.4.13)式.

有趣的是，标准化随机变数的分布函数与原来的参数无关. 故现在的 χ^2-变量，表面看来与 m_i,σ_i 有关，实际上它的密度函数也是与 m_i,σ_i 无关的，而只与自由度 n 有关. 这在实际应用上有着重要的意义.

证 由(2.4.9)式可得到此结果. ▮

推论 2

$$f_{\sqrt{\frac{\chi^2}{n}}}(x)=\frac{2\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}x^{n-1}e^{-\frac{nx^2}{2}}\ (x\geqslant 0), \qquad (2.4.16)$$

证 由于

$$P(\xi<x)=P(\xi^2<x^2) \qquad (设\ \xi>0, x>0),$$

得

$$F_{\xi}(x)=F_{\xi^2}(x^2),$$

从而

$$f_{\xi}(x)=f_{\xi^2}(x^2)\cdot 2x.$$

故

$$f_{\sqrt{\frac{\chi^2}{n}}}(x)=f_{\frac{\chi^2}{n}}(x^2)\cdot 2x.$$

即得(2.4.16). ▮

图 2.4.8 显示了 χ^2-分布之密度函数图.

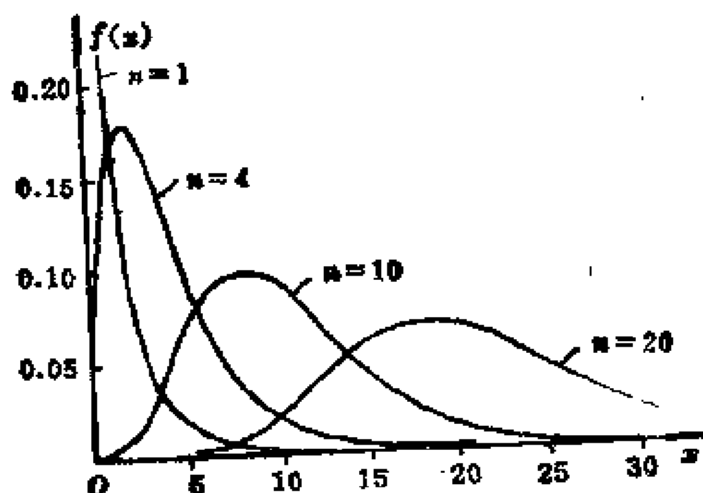


图 2.4.8 χ^2-分布密度函数图

定理 2.4.4(t-分布) 设 ξ, z 为相互独立随机变数, ξ 服从正态$N(0,1)$, z 服从自由度为 n 的χ^2-分布,则$t=\xi\Big/\sqrt{\frac{z}{n}}$ 的密

度函数为

$$f_t(x)=f_{\xi/\sqrt{\frac{z}{n}}}(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad (2.4.17)$$

称 $f_t(x)$ 是自由度为 n 的 t-分布(或 Student 分布) 的密度函数。

证 首先，易知 ξ 与 $\sqrt{z/n}$ 相互独立. 事实上，$F_{\xi,\sqrt{\frac{z}{n}}}(x,y)$

$=P\left(\xi<x,\ \sqrt{\frac{z}{n}}<y\right)=P(\xi<x,z<ny^2)=P(\xi<x)\cdot P(z<ny^2)$

$=P(\xi<x)\cdot P\left(\sqrt{\frac{z}{n}}<y\right)=F_\xi(x)\cdot F_{\sqrt{\frac{z}{n}}}(y)$，当 $y>0$ 时.

$F_{\xi,\sqrt{\frac{z}{n}}}(x,y)=0=F_\xi(x)\cdot 0=F_\xi(x)\cdot F_{\sqrt{\frac{z}{n}}}(y)$，当 $y\leqslant 0$ 时.

故得证 ξ 与 $\sqrt{z/n}$ 是相互独立的. 由(2.4.7)式，商的密度函数为

$$f_{\frac{\xi_1}{\xi_2}}(x)=\int_{-\infty}^{\infty} f_{\xi_1}(x_2x)f_{\xi_2}(x_2)\,|x_2|\,dx_2.$$

故

$$f_{\xi/\sqrt{\frac{z}{n}}}(x)=\frac{2\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\cdot\frac{1}{\sqrt{2\pi}}\int_0^{\infty}e^{-\frac{x^2x_2^2}{2}}x_2^n e^{-\frac{nx_2^2}{2}}dx_2,$$

令 $u=\frac{x_2\sqrt{n+x^2}}{\sqrt{2}}$，则上式变为

$$f_{\xi/\sqrt{\frac{z}{n}}}(x)=\frac{n^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}\,(n+x^2)^{\frac{n+1}{2}}}\int_0^{\infty}2u^ne^{-u^2}du$$

$$=\frac{n^{\frac{n}{2}}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}\,(n+x^2)^{\frac{n+1}{2}}}$$

$$=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

证明过程用到了公式

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx=2\int_0^\infty y^{2\alpha-1}e^{-y^2}dy \qquad (\alpha>0). \blacksquare$$

图 2.4.9 显示了 $n=2, n=5$ 时 t-分布与正态 $N(0,1)$二者密度之比较. 当 $n>30$ 时,二者相差无几,并且 $t(m)$, $t(n)$只要 m, $n>30$,也非常接近. 这可由附录 t-分布表直接看出.

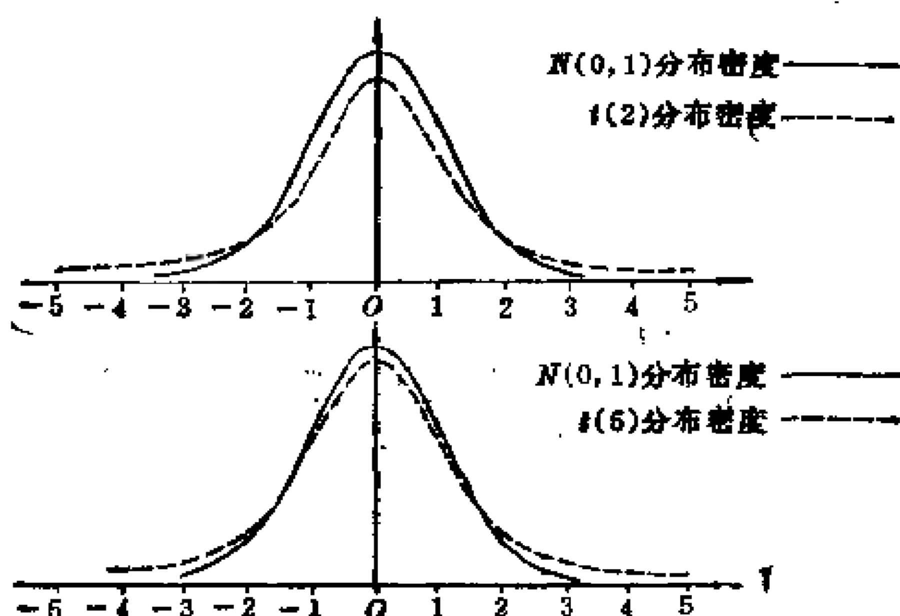


图 2.4.9 $n=2,5$ 时 t-分布密度与 $N(0,1)$正态分布密度之比较

当 $n=1$ 时,得到

$$f_t(x)=\frac{1}{\pi}\cdot\frac{1}{(1+x^2)}, \tag{2.4.18}$$

称之为哥西密度函数.

定理 2.4.5 (F-分布) 设 χ_m^2, χ_n^2 为独立随机变数, 分别具有自由度 m 及 n 的 χ^2-分布. 令 $\xi=\frac{\chi_m^2}{m}$, $\eta=\frac{\chi_n^2}{n}$, 则 $\zeta=\xi/\eta$ 的密度函数为

$$f_{\xi}(x)=\begin{cases}0, & \text{当} \quad x\leqslant 0;\\ \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & \text{当 } x>0.\end{cases} \tag{2.4.19}$$

称 $f_{\xi}(x)$ 为自由度为 m 及 n 的 F-分布的密度函数。

证 由商的密度函数公式(2.4.7)式:

$$f_{\frac{\xi_1}{\xi_2}}(x)=\int_{-\infty}^{\infty}f_{\xi_1}(x_2x)f_{\xi_2}(x_2)|x_2|dx_2,$$

及(2.4.15)式:

$$f_{\frac{\chi^2_n}{n}}(x)=\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{nx}{2}},\quad (x>0).$$

故得

$$\begin{aligned}f_{\frac{\chi^2_m}{m}\Big/\frac{\chi^2_n}{n}}(x)&=\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\cdot\frac{\left(\frac{m}{2}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{2}\right)}\int_0^{\infty}x_2{}^{\frac{m}{2}-1}x^{\frac{m}{2}-1}e^{-\frac{mxx_2}{2}}x_2^{\frac{n}{2}}e^{-\frac{nx_2}{2}}dx_2\\ &=Cx^{\frac{m}{2}-1}\int_0^{\infty}x_2{}^{\frac{m+n}{2}-1}e^{-\frac{1}{2}x_2(mx+n)}dx_2,\end{aligned}$$

其中

$$C=m^{\frac{m}{2}}n^{\frac{n}{2}}\Big/2^{\frac{m+n}{2}}\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right).$$

令 $u=\dfrac{x_2(mx+n)}{2}$,得

$$\begin{aligned}f_{\frac{\chi^2_m}{m}\Big/\frac{\chi^2_n}{n}}(x)&=C\cdot x^{\frac{m}{2}-1}\frac{2^{\frac{m+n}{2}}}{(mx+n)^{\frac{m+n}{2}}}\int_0^{\infty}u^{\frac{m+n}{2}-1}e^{-u}du\\ &=\frac{Cx^{\frac{m}{2}-1}\Gamma\left(\frac{m+n}{2}\right)}{2^{\frac{m+n}{2}}(mx+n)^{\frac{m+n}{2}}}\end{aligned}$$

$$=\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}\quad(\text{当 } x>0).\ \blacksquare$$

F-分布的密度函数图如图 2.4.10

在上面几个例子的计算中，我们可以看到，在计算随机变数的函数的分布函数(或密度函数)(2.4.2)式中，常常需要作积分变量替换去把它算出. 现对于连续型随机变数将这种计算过程总结成下面的定理.

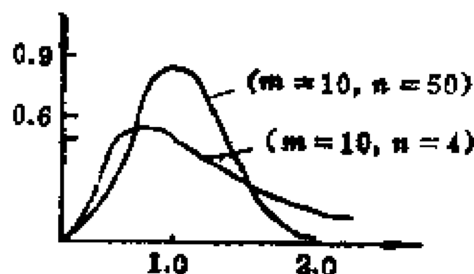


图 2.4.10 F-密度函数图

定理 2.4.6 设 ξ_1, ξ_2 的联合分布密度函数为 $f_\xi(x_1, x_2)$，若对于函数

$$\begin{cases}y_1=g_1(x_1,x_2),\\ y_2=g_2(x_1,x_2)\end{cases}\tag{2.4.20}$$

满足下述条件:

1° 存在唯一的反函数

$$\begin{cases}x_1=x_1(y_1,y_2),\\ x_2=x_2(y_1,y_2).\end{cases}\tag{2.4.21}$$

2° 有一切连续的一阶偏导数. 记

$$J=\begin{vmatrix}\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2}\\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2}\end{vmatrix}.$$

则 $\eta_1=g_1(\xi_1,\xi_2)$, $\eta_2=g_2(\xi_1,\xi_2)$ 的联合分布的密度函数为

$$f_{\xi_1,\xi_2}(x_1(y_1,y_2),x_2(y_1,y_2))\cdot|J|.\tag{2.4.22}$$

证 首先，由定理 2.3.2 可知，η_1, η_2 必为随机变数. 其次，η_1, η_2 的联合分布函数由(2.4.2)式，知为

$$F_{\eta_1,\eta_2}(y_1,y_2)=\iint_D f_{\xi_1,\xi_2}(x_1,x_2)dx_1dx_2$$
$$=\int_{-\infty}^{y_2}\int_{-\infty}^{y_1} f_{\xi_1,\xi_2}[x_1(v_1,v_2),x_2(v_1,v_2)]\cdot|J|dv_1dv_2.$$

故 η_1,η_2 的联合密度函数为

$$f_{\eta_1,\eta_2}(y_1,y_2)=f_{\xi_1,\xi_2}[x_1(y_1,y_2),x_2(y_1,y_2)]|J|. \quad ▌$$

若反函数(2.4.21)式不唯一，亦即一个 Y 平面的点 (y_1,y_2) 可对应于 X 平面上的多个点，则可想像为把 X 平面分割成许多不相交的部分 G_i，使每一部分中只有唯一的反函数，则在上述证明过程的积分变换中，就变为

$$\iint_D f_{\xi_1,\xi_2}(x_1,x_2)dx_1dx_2=\sum_{i=1}^{n}\iint_{G_i\cap D} f_{\xi_1,\xi_2}(x_1,x_2)dx_1dx_2$$
$$=\sum_{i=1}^{n}\int_{-\infty}^{y_2}\int_{-\infty}^{y_1} f_{\xi_1,\xi_2}(x_1^{(i)}(v_1,v_2),x_2^{(i)}(v_1,v_2))|J^{(i)}|dv_1dv_2$$
$$=\int_{-\infty}^{y_2}\int_{-\infty}^{y_1}\sum_{i=1}^{n} f_{\xi_1,\xi_2}(x_1^{(i)}(v_1,v_2),x_2^{(i)}(v_1,v_2))|J^{(i)}|dv_1dv_2.$$

因此，(η_1,η_2) 的密度函数为

$$f_{\eta_1,\eta_2}(y_1,y_2)=\sum_{i=1}^{n} f_{\xi_1,\xi_2}[x_1^{(i)}(y_1,y_2),x_2^{(i)}(y_1,y_2)]|J^{(i)}|.$$

若存在 E_0 使(2.4.20)对于 $(y_1,y_2)\in E_0$ 时，不能求出反函数，这就表明了二维随机变数 $(\eta_1,\eta_2)=(g_1(\xi_1,\xi_2),g_2(\xi_1,\xi_2))$ 没有可能于 E_0 中取值，故在 E_0 中之密度函数应为 0. 故最后可得 (η_1,η_2) 之密度函数的一般形式为

$$f_{\eta_1,\eta_2}(y_1,y_2)=\begin{cases}\sum_{i=1}^{n} f_{\xi_1,\xi_2}(x_1^{(i)}(y_1,y_2),x_2^{(i)}(y_1,y_2))|J^{(i)}| & \text{若(2.4.40)有解;}\\ 0, & \text{其它.}\end{cases}$$

下面举一例说明此定理的直接运用.

例 2.4.6 (瑞利(Rayleigh)分布). 设火炮射击时弹着点的坐标 (ξ_1,ξ_2) 服从下述的二维正态分布：

$$f_{\xi_1,\xi_2}(x_1,x_2)=\frac{1}{2\pi\sigma^2}e^{-\frac{x_1^2+x_2^2}{2\sigma^2}}, \tag{2.4.23}$$

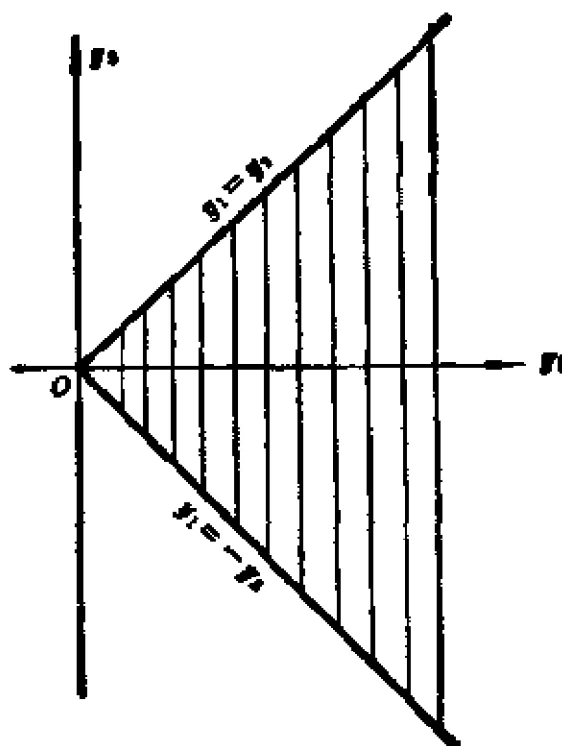


图 2.4.11

这里把目标作为坐标原点. 试求弹着点与目标之间的距离 $\eta_1=\sqrt{\xi_1^2+\xi_2^2}$ 所服从的分布密度函数 $f_{\eta_1}(x)$. (值得一提的是，这种分布叫做瑞利分布，在噪声理论和海浪理论中很有用.)

解 令

$$\begin{cases} y_1=\sqrt{x_1^2+x_2^2}, \\ y_2=x_2 \end{cases}$$

(注意 $y_1\geqslant|x_1|, y_1\geqslant|x_2|$). 解之，得

$$\begin{cases} x_1=\pm\sqrt{y_1^2-y_2^2} \\ x_2=y_2 \end{cases} \quad (y_1\geqslant|y_2|, y_2\in R_1),$$

$$J^{(1)}=\begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}=\begin{vmatrix} \frac{y_1}{\sqrt{y_1^2-y_2^2}} & \frac{-y_2}{\sqrt{y_1^2-y_2^2}} \\ 0 & 1 \end{vmatrix}=\frac{y_1}{\sqrt{y_1^2-y_2^2}},$$

$$J^{(2)} = \begin{vmatrix} \frac{-y_1}{\sqrt{y_1^2-y_2^2}} & \frac{y_2}{\sqrt{y_1^2-y_2^2}} \\ 0 & 1 \end{vmatrix} = -\frac{y_1}{\sqrt{y_1^2-y_2^2}}.$$

故在点(y_1, y_2)所属平面上，有

$$f_{\eta_1,\eta_2}(y_1, y_2) = \sum_{i=1}^{2} f_i(x_1^{(i)}(y_1, y_2), x_2^{(i)}(y_1, y_2)) |J^{(i)}|$$

$$= \frac{1}{2\pi\sigma^2} e^{-\frac{(y_1^2-y_2^2)+y_2^2}{2\sigma^2}} \cdot \frac{2y_1}{\sqrt{y_1^2-y_2^2}} = \frac{y_1 e^{-\frac{y_1^2}{2\sigma^2}}}{\pi\sigma^2\sqrt{y_1^2-y_2^2}},$$

$$f_{\eta_1}(y_1) = \int_{-\infty}^{\infty} f_{\eta_1,\eta_2}(y_1, y_2)\, dy_2 = \frac{y_1}{\pi\sigma^2} e^{-\frac{y_1^2}{2\sigma^2}} \int_{-y_1}^{y_1} \frac{dy_2}{\sqrt{y_1^2-y_2^2}}$$

$$= \frac{y_1}{\pi\sigma^2} e^{-\frac{y_1^2}{2\sigma^2}} \arcsin\frac{y_2}{y_1}\Bigg|_{y_2=-y_1}^{y_2=y_1} = \frac{y_1}{\sigma^2} e^{-\frac{y_1^2}{2\sigma^2}} \qquad (y_1>0).$$

最后可得

$$f_{\eta_1}(y_1) = \begin{cases} \frac{y_1}{\sigma^2} e^{-\frac{y_1^2}{2\sigma^2}}, & 当\ y_1>0; \\ 0, & 当\ y_1 \leqslant 0. \end{cases} \tag{2.4.24}$$ ▌

易知，瑞利分布是韦布分布在$m=2, \nu=0, x_0=2\sigma^2$之情形.

值得注意的是：定理 2.4.6 虽然具有普遍性，但对于有些具体问题，有时直接按定义去求出分布函数，再用求导法去得出密度函数，反而会更为简便. 例如，对于上例可直接简单计算如下：在$t \geqslant 0$情形，

$$F_{\eta_1}(t) = P(\sqrt{\xi_1^2+\xi_2^2} < t)$$

$$= \iint\limits_{\sqrt{x_1^2+x_2^2}<t} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x_1^2+x_2^2}{2\sigma^2}\right) dx_1 dx_2,$$

其中，记$\exp a = e^a$. 作极坐标变换

$$\begin{cases} x_1 = r\cos\theta \\ x_2 = r\sin\theta \end{cases}, \quad \frac{\partial(x_1, x_2)}{\partial(r, \theta)} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

故　当 $t\geqslant 0$ 时，

$$F_{\eta_1}(t)=\int_0^{2\pi}\int_0^t\frac{1}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)r\,dr\,d\theta$$

$$=\int_0^t\frac{1}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)r\,dr=1-e^{-\frac{t^2}{2\sigma^2}}. \tag{2.4.25}$$

从而得

$$f_{\eta_1}(t)=\frac{dF_{\eta_1}(t)}{dt}=\frac{t}{\sigma^2}e^{-\frac{t^2}{2\sigma^2}}，当\ t>0\ 时.$$

当 $t\leqslant 0$ 时，$f_{\eta_1}(t)=0$ 是显然的.

习　题

1. 甲、乙两名篮球队员独立地轮流投篮，直至某人投中篮圈为止. 今让甲先投，如果甲投中的概率为 0.4，乙为 0.6. 求各队员投篮次数的概率分布.

2. 画出下列随机变数的分布函数：

i)

ξ	0	$\frac{\pi}{2}$	π
$P\{\xi=x\}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

ii) $\eta=\frac{2}{3}\xi+2$，其中 ξ 同 i)；

iii) $\zeta=\cos\xi$，其中 ξ 与 i) 的相同.

3. 设随机变数 ξ 具有连续型分布：

$$F(x)=\int_{-\infty}^{x}f(t)\,dt,$$

试求下列随机变数的分布函数和密度函数：

i) $\eta=1/\xi$；ii) $\zeta=|\xi|$；iii) $U=e^{-\xi}$. 特别，若

$$f(x)=\begin{cases}2x，当\ 0<x\leqslant 1;\\ 0，其它.\end{cases}$$

求 i)、ii)、iii) 的密度函数.

4. 随机变数 ξ 的密度函数为

常用分布表

名称	概率函数或密度函数	图　　形	特征函数或母函数	数字特征①
退化分布（单点分布）	$P\{\xi=c\}=1$		$\varphi(t)=e^{jtc}$ $\psi(s)=s^c$ （c 为非负整数）	$E(\xi)=c$ $D(\xi)=0$
两点分布	$p_0=q,\quad p_1=p$ p,q非负，且$p+q=1$		$\varphi(t)=q+pe^{jt}$ $\psi(s)=q+ps$	$E(\xi)=p$ $D(\xi)=pq$
二项分布	$p_k=C_n^k p^k q^{n-k}$ $p>0,q>0$，且 $p+q=1$ $k=0,1,\cdots,n$.	二项分布 泊松分布	$\varphi(t)=(q+pe^{jt})^n$ $\psi(s)=(q+ps)^n$	$E(\xi)=np$ $D(\xi)=npq$
泊松分布	$p_k=\dfrac{\lambda^k}{k!}e^{-\lambda},\lambda>0$ $k=0,1,\cdots$		$\varphi(t)=e^{\lambda(e^{jt}-1)}$ $\psi(s)=e^{\lambda(s-1)}$	$E(\xi)=D(\xi)$ $=\lambda$
均匀分布	$f(x)=\begin{cases}\dfrac{1}{b-a}, & 当\ a\leqslant x\leqslant b\\ 0, & 其他\end{cases}$		$\varphi(t)=e^{j\left(\frac{a+b}{2}\right)t}\dfrac{\sin\left(\frac{b-a}{2}\right)t}{\left(\frac{b-a}{2}\right)t}$	$E(\xi)=\dfrac{b+a}{2}$ $D(\xi)=\dfrac{(b-a)^2}{12}$

续表

正态分布 $N(a,\sigma)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-a)^2}$ 常数 $\sigma>0$	[graph: $\frac{1}{\sigma\sqrt{2\pi}}$, a]	$\varphi(t)=e^{jat-\frac{1}{2}\sigma^2t^2}$	$E(\xi)=a$ $D(\xi)=\sigma^2$
指数分布	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & 当\ x\geqslant 0\\ 0, & 当\ x<0\end{cases}$ $\lambda>0$	[graph: $\lambda=0.5$; O, 1]	$\varphi(t)=\left(1-\frac{jt}{\lambda}\right)^{-1}$	$E(\xi)=\frac{1}{\lambda}$ $D(\xi)=\frac{1}{\lambda^2}$
韦布分布	$f(x)=\begin{cases}\frac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}} & 当x>\nu\\ 0, & 当x\leqslant\nu\end{cases}$ $m>0$, $x_0>0$, ν 均为常数	[graph: $x_0=1$, $\nu=0$; $m=3$, $m=2$; O, 1, 2; 1]		$E(\xi)=x_0^{\frac{1}{m}}\Gamma\left(\frac{1}{m}+1\right)+\nu$ $D(\xi)=x_0^{\frac{2}{m}}\left[\Gamma\left(\frac{2}{m}+1\right)-\Gamma^2\left(\frac{1}{m}+1\right)\right]$
Γ-分布	$f(x)=\begin{cases}\frac{x^a e^{-\frac{x}{\beta}}}{\beta^{a+1}\Gamma(a+1)}, & x\geqslant 0\\ 0, & x<0\end{cases}$ $\beta>0$, $a>-1$	[graph: $a=0$, $a=1$, $a=3$; $3/4$, $1/2$, $1/4$; O, 1, 2, 3, 4]	$\varphi(t)=\frac{1}{(1-j\beta t)^{a+1}}$	$E(\xi)=\beta(a+1)$ $D(\xi)=\beta^2(a+1)$

名称	概率函数或密度函数	图　　形	特征函数或母函数	数字特征
对数正态分布	$f(x)=\begin{cases}\dfrac{\lg e}{\sqrt{2\pi}\sigma x}e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2}, & 当 x>0\\ 0, & 当 x\leqslant 0\end{cases}$	$\sigma=0.1$; $\sigma=0.3$; $\sigma=0.5$; O		$E(\xi)=10^{\mu+\frac{\sigma^2\ln 10}{2}}$ $D(\xi)=10^{2\mu+\sigma^2\ln 10}\cdot(10^{\sigma^2\ln 10}-1)$
哥西分布	$f(x)=\dfrac{1}{\pi}\dfrac{\lambda}{\lambda^2+(x-\mu)^2}$	$\frac{1}{\pi\lambda}$; μ	$\varphi(t)=e^{i\mu t-\lambda\|t\|}$	各阶矩都不存在
三角形分布	$f(x)=\begin{cases}\dfrac{1}{4}(2+x), & 当 -2\leqslant x<0\\ \dfrac{1}{4}(2-x), & 当 0\leqslant x\leqslant 2\\ 0, & 当 \|x\|>2\end{cases}$	$f(x)$; $\frac{1}{2}$; -2; O; 2; x	$\varphi(t)=\left(\dfrac{\sin t}{t}\right)^2$	$E(\xi)=0$ $D(\xi)=\dfrac{2}{3}$

χ^2-分布	$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & 当x\geqslant 0\\ 0, & 当x<0\end{cases}$ n为正整数	0.5, 0.2, 0.1; 0, 2, 4, 6; $n=1$, $n=2$, $n=6$	$\varphi(t)=\dfrac{1}{(1-2jt)^{\frac{n}{2}}}$	$E(\xi)=n$ $D(\xi)=2n$
t-分布	$f(x)=\dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-\frac{n+1}{2}}$ n为正整数	0.4; −4, −2, 0, 2, 4 $n=3$时		$E(\xi)=0$ ($n>1$时) $D(\xi)=\dfrac{n}{n-2}$ ($n>2$时)
F-分布	$f(x)=\begin{cases}\dfrac{\Gamma\left(\frac{m+n}{2}\right)m^{\frac{m}{2}}n^{\frac{n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\cdot\dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & 当x>0\\ 0, & 当x\leqslant 0\end{cases}$	0.9, 0.6; 0, 1.0, 2.0; ($m=10$, $n=50$), ($m=10$, $n=4$)		$E(\xi)=\dfrac{n}{n-2}$ ($n>2$时) $D(\xi)=\dfrac{2n^2(n+m-2)}{m(n-2)^2(n-4)}$, (当$n>4$时)

二维常用分布表[1]

名称	概率函数或密度函数	特征函数或母函数	数字特征
二维两点分布	$p(i,k)=\begin{cases} q, & 当\ i=0, \quad k=0 \\ p, & 当\ i=1 \quad k=1 \\ 0, & 其他 \end{cases}$	$\varphi(t_1,t_2)=q+pe^{i(t_1+t_2)}$ $\psi(s_1,s_2)=q+ps_1s_2$	$E(\xi)=E(\eta)=p$ $b_{ik}=pq$ $\begin{pmatrix} i=0,1 \\ k=0,1 \end{pmatrix}$
二维均匀分布	$f(x,y)=\begin{cases} \frac{1}{(b_1-a_1)(b_2-a_2)}, & 当\ a_1\leqslant x\leqslant b_1,\ a_2\leqslant y\leqslant b_2 \\ 0, & 其他 \end{cases}$	$\varphi(t_1,t_2)=\frac{e^{i\left(\frac{b_1+a_1}{2}t_1+\frac{b_2+a_2}{2}t_2\right)}\sin\left(\frac{b_1-a_1}{2}\right)t_1\sin\left(\frac{b_2-a_2}{2}\right)t_2}{\left(\frac{b_1-a_1}{2}\right)\left(\frac{b_2-a_2}{2}\right)t_1t_2}$	$E(\xi)=\frac{b_1+a_1}{2}$ $E(\eta)=\frac{b_2+a_2}{2}$ $b_{ik}=\begin{cases} \frac{(b_1-a_1)^2}{12} & 当\ i=k=1 \\ \frac{(b_2-a_2)^2}{12} & 当\ i=k=2 \\ 0 & 其他 \end{cases}$
二维正态分布	$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}$ $\times e^{-\frac{1}{2(1-r^2)}\left[\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right]}$	$\varphi(t_1,t_2)=e^{i(m_1t_1+m_2t_2)-\frac{1}{2}(\sigma_1^2t_1^2+2r\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2)}$	$E(\xi)=m_1$ $E(\eta)=m_2$ $D(\xi)=\sigma_1^2$ $D(\eta)=\sigma_2^2$ $\rho=r$

① 表中出现的数字特征与特征函数、母函数等概念的定义分别见第三、四两章。

$$f(x)=\begin{cases} a\cos x, & \text{当}-\frac{\pi}{2}\leqslant x<\frac{\pi}{2}; \\ 0, & \text{其它}. \end{cases}$$

i) 求系数 a; ii) 作出 $f(x)$ 及 $F(x)$ 的图形;

iii) 求 $P\left(0\leqslant\xi\leqslant\frac{\pi}{4}\right)$.

5. 问 A 为何值时, $F(x)=A-e^{-x}$ $(0\leqslant x<\infty)$ 是一随机变数 ξ 的分布函数(设当 $x<0$ 时, $F(x)=0$)?

6. 设 ξ 是 $[0,1]$ 上的连续型随机变数, 且 $P(\xi\leqslant 0.29)=0.75$, 如果 $\eta=1-\xi$, 试决定 k, 使得 $P(\eta\leqslant k)=0.25$.

7. 设某动物生下 r 个蛋的概率是 $P(\xi=r)=\frac{\lambda^r}{r!}e^{-\lambda}$. 若每一个蛋能发育成小动物的概率是 p, 且各个蛋能否发育成小动物是彼此独立的. 证明恰有 k 个后代的概率分布是具有参数为 λP 的泊松分布.

8. 在 $(0,a)$ 线段上任意抛两个点(抛掷的二点的位置在 $(0, a)$ 上独立地服从均匀分布). 试求两点间距离的分布函数.

9. 设随机变数 ξ 具有严格单调上升连续的分布函数 $F(x)$. 求 $\eta=F(\xi)$ 的分布函数.

10. 设 ζ 是在任何有限区间 (a, b) 上均有 $P\{\zeta\in(a, b)\}>0$ 的连续型随机变数. 其分布函数为 F_ζ. 如果 ξ 在 $[0,1]$ 上服从均匀分布, 令 $\eta=F_\zeta^{-1}(\xi)$ 证明 η 具有与 ζ 相同的分布函数 F_ζ.

11. 设 $F_1(x)$、$F_2(x)$ 为两个分布函数, 问: i) $F_1(x)+F_2(x)$ 是否为分布函数? ii) 若 $a_1>0, a_2>0$ 均为常数, 且 $a_1+a_2=1$. 证明

$$a_1F_1(x)+a_2F_2(x)$$

为分布函数.

12. 证明任何分布函数具有下列性质:

$$\lim_{x\to+\infty}x\int_x^{\infty}\frac{1}{z}dF(z)=0; \qquad \lim_{x\to 0^+}x\int_x^{\infty}\frac{1}{z}dF(z)=0;$$

$$\lim_{x\to-\infty}x\int_{-\infty}^{x}\frac{1}{z}dF(z)=0; \qquad \lim_{x\to 0^-}x\int_{-\infty}^{x}\frac{1}{z}dF(z)=0.$$

13. 求证: 如果 $F(x)$ 是分布函数, 则对任何 $h\neq 0$, 函数

$$\Phi(x)=\frac{1}{h}\int_x^{x+h}F(y)\,dy; \quad \psi(x)=\frac{1}{2h}\int_{x-h}^{x+h}F(y)\,dy$$

也是分布函数.

14. 设二维随机变数(ξ,η)在以原点为中心，r为半径的圆上服从均匀分布，求联合密度函数及各边沿分布密度函数.

15. 设二维随机变数(ξ,η)的密度函数为

$$f(x,y)=\begin{cases}x^2+\dfrac{xy}{3}, & 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 2;\\ 0, & \text{其它}.\end{cases}$$

求：i) (ξ,η)的边沿分布密度函数；

ii) ξ,η的条件分布密度函数；

iii) $P(\xi+\eta>1)$；$P(\eta<\xi)$及$P\left(\eta<\dfrac{1}{2}\,\middle|\,\xi<\dfrac{1}{2}\right)$.

16. 设二维随机变数(ξ_1,ξ_2)的密度函数是

$$\frac{1}{\Gamma(k_1)\Gamma(k_2)}x_1^{k_1-1}(x_2-x_1)^{k_2-1}e^{-x_2},$$

求ξ_1和ξ_2的边沿分布密度函数.

17. 设随机变数ξ_1: $P(\xi_1=0)=1$，而ξ_2是任意的随机变数. 证明：ξ_1与ξ_2相互独立.

18. 设(ξ,η)具有下述联合分布密度函数，问ξ与η是否相互独立?

i)

$$f(x,y)=\begin{cases}xe^{-x}\dfrac{1}{(1+y)^2}, & x>0,\ y>0;\\ 0, & \text{其它},\end{cases}$$

ii) $f(x,y)=8xy,\ 0\leqslant x\leqslant y\leqslant 1$.

19. 设随机变数(ξ,η)服从二维正态分布

$$f(x,y)=\frac{1}{2\pi ab}e^{-\frac{1}{2}\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)},$$

求(ξ,η)取值于椭圆：$\dfrac{x^2}{a^2}+\dfrac{y^2}{b^2}=r^2$内的概率.

20. 设随机变数ξ的密度函数为

$$f(x)=Ae^{-|x|}\quad(-\infty<x<\infty).$$

求：i) 系数A；　　ii) $P(0\leqslant\xi\leqslant 1)$；

iii) 分布函数$F(x)$.

21. i) 设ξ是$[0,\pi]$上的均匀分布，求$\eta=\sin\xi$的分布函数.

ii) 设ξ是$\left[-\dfrac{\pi}{2},\ \dfrac{\pi}{2}\right]$上的均匀分布，求$\eta=\cos\xi$的分布函数.

22. 对事件 $A_i(i=1,2,\cdots,n)$ 定义随机变数

$$\xi_i=\begin{cases}0, & 若\ \omega\bar{\in}A_i;\\ 1, & 若\ \omega\in A_i.\end{cases}$$

试证：事件 $A_1,A_2,\cdots,A_n$ 相互独立的充要条件是 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立.

23. 设 ξ_1 与 ξ_2 相互独立，并具有共同的几何分布 $P\{\xi_i=k\}=pq^k$，$(i=1,2;k=0,1,2,\cdots)$.

i) 证明：

$$P(\xi_1=k|\xi_1+\xi_2=n)=\frac{1}{n+1}\quad(k=0,1,\cdots,n);$$

ii) 求 $\eta=\max\{\xi_1,\xi_2\}$ 的分布；

iii) 求 η 与 ξ_1 的联合分布.

24. 设随机变数 ξ,η 相互独立，且都服从泊松分布：

$$f_\xi(m)=\frac{\lambda_1^m}{m!}e^{-\lambda_1},\quad m=0,1,2,\cdots;$$

$$f_\eta(n)=\frac{\lambda_2^n}{n!}e^{-\lambda_2},\quad n=0,1,2,\cdots.$$

求证：$\xi+\eta$ 也服从泊松分布，并且对于给定的 $\xi+\eta$，ξ 的条件分布是二项分布：

$$P(\xi=k|\xi+\eta=N)=b\left(k,N,\frac{\lambda_1}{\lambda_1+\lambda_2}\right).$$

25. 设 ξ 与 η 相互独立，且有下述分布：求 $\xi+\eta$ 的分布.

i) (ξ,η) 服从正态分布：

$$f(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)};$$

ii) ξ 与 η 都服从二项分布：

$$P_\xi(m)=C_2^m\left(\frac{1}{2}\right)^m\left(\frac{1}{2}\right)^{2-m}\quad(m=0,1,2),$$

$$P_\eta(n)=C_2^n\left(\frac{1}{3}\right)^n\left(\frac{2}{3}\right)^{2-n}\quad(n=0,1,2);$$

iii) ξ 在 $[0,1]$ 内服从均匀分布，η 在 $[0,2]$ 上服从辛普生分布：

$$f_\eta(y)=\begin{cases}y, & y\in[0,1],\\ 2-y, & y\in[1,2],\\ 0, & 其它;\end{cases}$$

iv) ξ,η 分别为 $(-5,1)$ 与 $(1,5)$ 内的均匀分布；

v) ξ 服从 $N(a,\sigma)$，η 为 $[-b,b]$ 上的均匀分布；

vi) ξ,η 的密度函数分别为

$$f_\xi(x)=\frac{1}{2}e^{-\frac{x}{2}},\quad 0\leqslant x<\infty$$

$$f_\eta(y)=\frac{1}{3}e^{-\frac{y}{3}},\quad 0\leqslant y<\infty$$

vii) 设随机变数 ξ,η 相互独立，ξ 在 $[-h,h]$ 上均匀分布，η 有分布函数 $F_\eta(y)$.

26. 求证：如果随机变数 ξ 与 η 独立同分布：

$$f_\xi(x)=f_\eta(x)=\begin{cases}e^{-x}, & x>0;\\ 0, & x\leqslant 0,\end{cases}$$

则

i) $\xi+\eta$ 与 ξ/η;

ii) $\xi+\eta$ 与 $\xi/(\xi+\eta)$

也是相互独立的.

27. 如果 ξ 与 η 相互独立，均服从 $N(0,1)$，则 $\xi^2+\eta^2$ 与 ξ/η 相互独立.

28. 设连续型随机变数 ξ,η 相互独立，分别对下面三种情形求 $\xi\eta$ 以及 $\xi-\eta$ 的分布函数：

i) 分别有分布函数 $F_1(x),F_2(x)$;

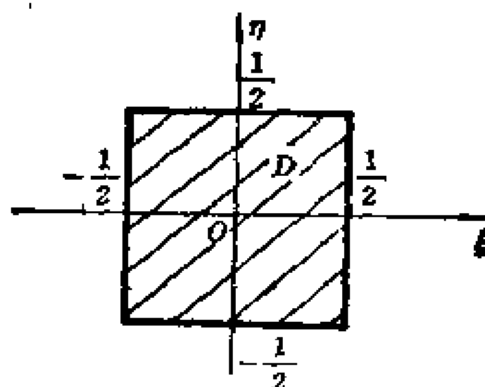
ii) 均匀分布于 $(-a,a)$ 内；

iii) 服从 $N(0,1)$.

29. 设二维随机变数 (ξ,η) 服从均匀分布：

$$f(x,y)=\begin{cases}1, & (x,y)\in D;\\ 0, & \text{其它},\end{cases}$$

其中 D 为右图所示的阴影区域. 求随机变数 $\zeta=\xi\eta$ 的密度函数.



第 29 题图

30. 设二维随机变数 (ξ,η) 服从正态分布

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2}+\frac{y^2}{\sigma_2^2}\right)}.$$

问随机变数 $\xi-\eta$ 服从什么分布？

31. 设随机变数 ξ,η 相互独立，分别对下面二种情形，求 ξ/η 的分布函数：

i）皆服从 $N(0,1)$ 分布；

ii）皆服从 $(0,a)$ 上的均匀分布；

32. 设 ξ，η 相互独立，它们的密度函数恒不为零，且二次可微．试证：若 $\xi+\eta$，$\xi-\eta$ 相互独立，则 ξ，η 从而 $\xi+\eta$，$\xi-\eta$ 都服从正态分布．

33. 对任一大学生，他的生日在一年中任一天的概率均为 1/365，若某一学校有 730 名大学生，问有 4 名大学生的生日为元旦的概率是多少？

34. 设某车间有 200 台同一型号的车床．由于种种原因，每台车床时常需要停车．假定各台车床的停车或开动是相互独立的，且每台车床有 60% 的时间开动，开动时需要消耗的电能为 E．问至少要供给这个车间多少电能，才能以 99.9% 的概率保证这个车间不致因为供电不足而影响生产．

第三章　随机变数的数字特征

所谓随机变数的数字特征，是指连系于它的分布函数的某些数，如平均值，最大可能值等．它们反映随机变数的某方面的特征．在第二章我们举出常见的随机变数分布函数的各种例子，很多分布函数含有一个或多于一个参数(如泊松分布含有一个参数λ，正态分布含有两个参数a和σ)，这些参数往往是由某些数字特征或其他数值所决定的，因此找到这些特征，分布函数(或概率函数，密度函数)跟着就确定了．但对一般随机变数，要完全确定它的分布函数就不那么容易了，不过在许多实际问题中，我们并不需要完全知道分布函数，我们只需要知道随机变数的某些特征也就够了．例如，在测量某物件的长度时，测量的结果是一随机变数．在实际工作中，往往用测量长度的平均数来代表这一物件的长度．又如对一射手的技术评定，除了要了解命中环数的平均值，同时还必须考虑稳定情况，命中点分散还是比较集中？这些特征往往为数字特征所决定．由此可见，随机变数的数字特征的研究具有理论上和实际上的重要意义．

§3.1　数学期望与方差

先看几个例子．

例3.1.1　设射击手甲与乙在同样条件下进行射击，其命中的环数是一随机变数．假设由历史纪录可得到它们分别有下面的分布列．

$\xi_甲$	10	9	8	7	6	5	0
P	0.5	0.2	0.1	0.1	0.05	0.05	0

$\xi_乙$	10	9	8	7	6	5	0
P	0.1	0.1	0.1	0.1	0.2	0.2	0.2

(其中 0 环表示脱靶). 试问,应如何来评定甲、乙的技术优劣.

解 由射手甲的分布列很清楚地知道,他命中 10 环的概率是 0.5. 换句话说,他发出 100 粒子弹,约有 50 粒子弹命中 10 环. 同理,约有 20 粒命中 9 环,约有 10 粒命中 8 环和 7 环,约有 5 粒命中 6 环和 5 环,没有脱靶的. 这样"平均"起来甲命中环数约为

$$\frac{1}{100}[10\times 50+9\times 20+8\times 10+7\times 10+6\times 5+5\times 5+0\times 0]$$

$$=8.85(\text{环}).$$

我们把它记为 $E(\xi_甲)$,对上式稍作变化得

$$\begin{aligned}E(\xi_甲)&=10\times\frac{50}{100}+9\times\frac{20}{100}+8\times\frac{10}{100}+7\times\frac{10}{100}\\&\quad+6\times\frac{5}{100}+5\times\frac{5}{100}+0\times\frac{0}{100}\\&=10\times 0.5+9\times 0.2+8\times 0.1+7\times 0.1+6\times 0.05\\&\quad+5\times 0.05+0\times 0\\&=8.85(\text{环}).\end{aligned}\tag{3.1.1}$$

(3.1.1)式可以作为射手甲击中环数的理论平均值. 因为它是由 $\xi_甲$ 的理论取值与理论取值的概率相乘后求和得到(亦即加权平均得到)的.

同样,对于射手乙理论平均命中环数为

$$\begin{aligned}E(\xi_乙)&=10\times 0.1+9\times 0.1+8\times 0.1+7\times 0.1\\&\quad+6\times 0.2+5\times 0.2+0\times 0.2\\&=5.6(\text{环}).\end{aligned}\tag{3.1.2}$$

由(3.1.1)及(3.1.2)看到，从理论平均命中环数看，射手甲的射击水平高于射手乙的射击水平. 同时，我们也看到，这种反映随机变数取值"平均"意义特性的数值，恰好是这个随机变数取的一切可能值与相应概率乘积的总和. 即若随机变数 ξ 取值为 x_1, x_2, $\cdots$，取这些值相应的概率为 $p_1, p_2, \cdots$，则反映 ξ"平均"意义的数字特征为 $E(\xi)=x_1p_1+x_2p_2+\cdots=\sum\limits_i x_ip_i$，并把它叫做 ξ 的平均值.

对于射手甲、乙的技术水平，除了上述从平均值的角度来考虑外，还可从射击命中环数的集中或离散程度来考虑. 由上所述，射手甲命中环数的平均值是 8.85. 因此，他命中 10 环与平均值 8.85 的偏离值为 $10-8.85=1.15$，偏离的平方值为 $(10-8.85)^2$. 但射手甲命中 10 环的概率为 0.5，因而，在射击 100 发子弹中约有 50 次出现偏离的平方为 $(10-8.85)^2$. 同样理由，可得下表(3.1.1).

表 3.1.1

偏离值	$10-8.85$	$9-8.85$	…	$5-8.85$	$0-8.85$
偏离值的平方	$(10-8.85)^2$	$(9-8.85)^2$	…	$(5-8.85)^2$	$(0-8.85)^2$
概率	0.5	0.2	…	0.05	0

按平均值 $E(\xi_甲)$ 的想法，射手甲射击的"平均"的平方偏差值可为

$$\frac{1}{100}[(10-8.85)^2\times 50+(9-8.85)^2\times 20+\cdots+(5-8.85)^2\times 5+(0-8.85)^2\times 0],$$

并记它为 $D(\xi_甲)$，改写后为

$$D(\xi_甲)=(10-8.85)^2\times 0.5+(9-8.85)^2\times 0.2+\cdots+(5-8.85)^2\times 0.05+(0-8.85)^2\times 0=2.2275. \quad (3.1.3)$$

同理，可得

$$D(\xi_乙)=(10-5.6)^2\times0.1+(9-5.6)^2\times0.1+\cdots$$
$$+(5-5.6)^2\times0.2+(0-5.6)^2\times0.2$$
$$=10.24. \tag{3.1.4}$$

比较(3.1.3)及(3.1.4)两式得知，从平方偏离值的"平均"值看，射手甲的技术优于射手乙.

把(3.1.3)与(3.1.4)抽象成一般形式，得到：若离散型随机变数ξ取值为$x_1,x_2,\cdots$，相应的概率为$p_1,p_2,\cdots$，则反映ξ"平均"平方偏离值特性的数值为

$$[x_1-E(\xi)]^2p_1+[x_2-E(\xi)]^2p_2+\cdots=\sum_i[x_i-E(\xi)]^2p_i.$$

记它为$D(\xi)$.

这里，我们求平方偏离值的"平均"值，而不去求偏离值的"平均"值，原因在于：偏离值有正，有负，在相加的过程中，不应让它们互相抵消，而应让每一次偏离值(不管是正是负)都被考虑进去，故可考虑偏离值的平方值，乘以相应的概率并加以求和.

一、离散型和连续型随机变数的数学期望和方差

下面，我们引进它们的正式定义：

定义 3.1.1 设ξ为离散型随机变数，其分布列由下表给出.

ξ	x_0	x_1	x_2	$\cdots$	x_n	$\cdots$
$P(\xi=x_i)$	p_0	p_1	p_2	$\cdots$	p_n	$\cdots$

其中

$$p_i\geqslant0,\quad \sum_{i=0}^{\infty}p_i=1.$$

若

$$\sum_{i=0}^{\infty}|x_i|p_i<\infty ①,$$

① 我们规定级数绝对收敛，是为了保证即使任意调乱求和的次序，也不会影响级数的收敛与和数，而这在直观上是关于"平均值"的一个很自然的合理要求.

记

$$E(\xi)=\sum_{i=0}^{\infty}x_i p_i, \tag{3.1.5}$$

称 $E(\xi)$ 为 ξ 的数学期望①.

若

$$\sum_i[x_i-E(\xi)]^2p_i<\infty,$$

记

$$D(\xi)=\sum_i[x_i-E(\xi)]^2p_i, \tag{3.1.6}$$

称 $D(\xi)$ 为 ξ 的方差.

由 $D(\xi)$ 的定义，立即可得:

$$D(\xi)=\sum_i x_i^2 p_i-[E(\xi)]^2. \tag{3.1.7}$$

事实上,

$$\begin{aligned}D(\xi)&=\sum_i[x_i-E(\xi)]^2p_i\\&=\sum_i\{x_i^2-2x_iE(\xi)+[E(\xi)]^2\}p_i\\&=\sum_i x_i^2p_i-2E(\xi)\sum_i x_ip_i+[E(\xi)]^2\sum_i p_i\\&=\sum_i x_i^2p_i-[E(\xi)]^2.\end{aligned}$$

如上所述，数学期望刻划随机变数取值的"平均数"，而方差则刻划了该随机变数围绕"平均数"的离散程度. 我们还可以给予物理解释，加深我们对它们的理解. 例如在(3.1.5)式中，若把 x_i 想像为第 i 个质点所处位置的横坐标，p_i 表示第 i 个质点的质量，则(3.1.5)式表示质点系

$$\begin{pmatrix}x_0 & x_1 & \cdots x_n & \cdots\\ p_0 & p_1 & \cdots p_n & \cdots\end{pmatrix}$$

---

① 符号 $E(\xi)$ 有时简写为 $E\xi$. 同样，$E(\xi^k)$ 有时简写为 $E\xi^k$ ($E(\xi^k)$ 定义见后面 §3.2). 对于连续型随机变数也这样规定.

的重心横坐标，而(3.1.6)则表示该质点系相对于通过重心 $E(\xi)$ 的纵轴的转动惯量.

下面求几种常见的随机变数的数学期望与方差.

设随机变数 ξ 服从退化分布，即 $P\{\xi=c\}=1$，其中 c 为常数。显然有 $E(\xi)=c, D(\xi)=0$.

例 3.1.2 （两点分布）

设 ξ 的分布列由下表给出

ξ	0	1
P_i	$1-p$	p

求其数学期望与方差.

解 $E(\xi)=0\cdot(1-p)+1\cdot p=p$. 若记 $q=1-p$，则

$$D(\xi)=E[\xi-E(\xi)]^2=(0-p)^2\cdot(1-p)+(1-p)^2\cdot p$$

$$=p(1-p)=pq. \tag{3.1.8}$$

例 3.1.3 （二项分布） 若 ξ 的分布列由下表给出

ξ	0	1	…	k	…	n
P_i	$(1-p)^n$	$C_n^1 p(1-p)^{n-1}$	…	$C_n^k p^k(1-p)^{n-k}$	…	p^n

其中 $0\leqslant p\leqslant 1, 1-p=q$ 为常数，求数学期望与方差.

解 $$E(\xi)=\sum_{k=0}^{n}k\cdot C_n^k p^k(1-p)^{n-k}=np\sum_{k=1}^{n}C_{n-1}^{k-1}p^{k-1}(1-p)^{n-k}$$

$$=np\sum_{k=0}^{n-1}C_{n-1}^{k}p^k(1-p)^{n-1-k}=np[p+(1-p)]^{n-1}$$

$$=np. \tag{3.1.9}$$

由(3.1.7)式得

$$D(\xi)=\sum_{k=0}^{n}k^2C_n^kp^k(1-p)^{n-k}-[E(\xi)]^2$$

$$=np\sum_{k=1}^{n}k\cdot C_{n-1}^{k-1}p^{k-1}(1-p)^{n-k}-(np)^2$$

$$=np\sum_{k=0}^{n-1}(k+1)C_{n-1}^{k}p^k(1-p)^{n-1-k}-(np)^2$$

$$=np\left\{\sum_{k=0}^{n-1}kC_{n-1}^{k}p^k(1-p)^{n-1-k}+[p+(1-p)]^{n-1}\right\}-(np)^2$$

$$=np[(n-1)p+1]-(np)^2=n\cdot p\cdot(1-p)$$

$$=npq. \tag{3.1.10}$$

由于二项分布被两参数 n 和 p 所决定(亦即，n,p 给定后，二项分布唯一地确定)，而由(3.1.9)与(3.1.10)又可知，当 $E(\xi)$ 及 $D(\xi)$ 之值给定后，n,p 便唯一地确定. 故可知服从二项分布的随机变数的分布列，由它的数学期望与方差所唯一确定.

例3.1.4 (泊松分布) 设随机变数 ξ 的分布列由下式给出.

ξ	0	1	…	k	…
$P(\xi=k)$	$e^{-\lambda}$	$e^{-\lambda}\frac{\lambda}{1!}$	…	$e^{-\lambda}\frac{\lambda^k}{k!}$	…

试求 $E(\xi)$ 与 $D(\xi)$.

解 $E(\xi)=\sum_{k=0}^{\infty}k\cdot e^{-\lambda}\cdot\left(\frac{\lambda^k}{k!}\right)=\lambda\sum_{k=1}^{\infty}e^{-\lambda}\left(\frac{\lambda^{k-1}}{(k-1)!}\right)=\lambda.$

由(3.1.7)得

$$D(\xi)=\sum_{k=0}^{\infty}k^2e^{-\lambda}\left(\frac{\lambda^k}{k!}\right)-\lambda^2=\lambda\sum_{k=1}^{\infty}k\cdot e^{-\lambda}\left(\frac{\lambda^{k-1}}{(k-1)!}\right)-\lambda^2$$

$$=\lambda\sum_{k=0}^{\infty}(k+1)e^{-\lambda}\frac{\lambda^k}{k!}-\lambda^2=\lambda[\lambda+1]-\lambda^2=\lambda.$$

这说明对于服从泊松分布的随机变数而言，它的数学期望与方差

相等. 即

$$E(\xi)=D(\xi)=\lambda. \tag{3.1.11}$$

因而得到:服从泊松分布的随机变数的分布列由它的数学期望(方差)唯一确定.

在讨论连续型随机变数的数字特征之前，让我们再来计算一下超几何分布的期望与方差. 回忆起在第二章 §1 中，称 ξ 服从超几何分布,若

ξ	0	1	…	k	…	l
$P(\xi=k)$	$\frac{C_M^0 C_{N-M}^n}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	…	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	…	$\frac{C_M^l C_{N-M}^{n-l}}{C_N^n}$

其中 $0\leqslant n\leqslant N, 0\leqslant M\leqslant N, l=\mathrm{Min}(M,n)$，并且它们都是非负整数. 现求 $E(\xi)$ 与 $D(\xi)$. 有

$$E(\xi)=\sum_{k=0}^{l} kC_M^k C_{N-M}^{n-k}/C_N^n=\sum_{k=1}^{l}\frac{M!}{(k-1)!(M-k)!}\cdot\frac{C_{N-M}^{n-k}}{C_N^n}$$

$$=M\cdot\sum_{k=1}^{l}C_{M-1}^{k-1}\cdot C_{N-1-(M-1)}^{n-1-(k-1)}\Big/\left(C_{N-1}^{n-1}\cdot\frac{N}{n}\right)$$

$$=n\cdot M/N.$$

$$D(\xi)=\sum_{k=0}^{l}k^2C_M^kC_{N-M}^{n-k}/C_N^n-n^2M^2/N^2$$

$$=\sum_{k=1}^{l}k^2\cdot P(N,M;n,k)-n^2M^2/N^2$$

$$\left(\text{其中记 } P(N,M;n,k)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}\right)$$

$$=\frac{nM}{N}\sum_{k=1}^{l}kP(N-1,M-1;n-1,k-1)-n^2M^2/N^2$$

$$=\frac{nM}{N}\sum_{k=1}^{l}(k-1)P(N-1,M-1;n-1,k-1)$$

$$+\frac{nM}{N}-\frac{n^2M^2}{N^2}$$

$$=\frac{nM}{N}\cdot\left(\frac{(n-1)(M-1)}{N-1}+1-\frac{nM}{N}\right)$$

$$=\frac{nM}{N}\cdot\frac{(N-n)(N-M)}{N(N-1)}=\frac{nM(N-n)(N-M)}{N^2(N-1)}.$$

下面讨论连续型随机变数的期望与方差.

对于连续型随机变数,仿照定义 3.1.1 有

定义 3.1.2 设 ξ 为具有密度函数 $f(x)$的随机变数.

若

$$\int_{-\infty}^{\infty}|x|f(x)dx<\infty,$$

记

$$E(\xi)=\int_{-\infty}^{\infty}xf(x)dx, \tag{3.1.12}$$

称 $E(\xi)$为 ξ 的数学期望.

若

$$\int_{-\infty}^{\infty}[x-E(\xi)]^2f(x)dx<\infty,$$

记

$$D(\xi)=\int_{-\infty}^{\infty}[x-E(\xi)]^2f(x)dx,$$

称 $D(\xi)$为 ξ 的方差.

同离散型随机变数一样,有

$$D(\xi)=\int_{-\infty}^{\infty}x^2f(x)dx-[E(\xi)]^2. \tag{3.1.13}$$

事实上

$$D(\xi)=\int_{-\infty}^{\infty}[x-E(\xi)]^2f(x)dx$$

$$=\int_{-\infty}^{\infty}x^2f(x)dx$$

$$-2E(\xi)\int_{-\infty}^{\infty}xf(x)dx+[E(\xi)]^2\int_{-\infty}^{\infty}f(x)dx$$

$$=\int_{-\infty}^{\infty}x^2f(x)dx-[E(\xi)]^2.$$

可以对数学期望与方差作如下物理解释:

数学期望与方差可分别看成一个具有密度为 $f(x)$ 的质量连续分布于整个横坐标轴上的连续质点系的重心和转动惯量。

例 3.1.5(均匀分布)

设随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & 当 a\leqslant x\leqslant b;\\ 0, & 其它,\end{cases}$$

试求 $E(\xi)$ 与 $D(\xi)$。

解 $E(\xi)=\int_{-\infty}^{\infty}xf(x)dx=\int_a^b\dfrac{x}{b-a}dx=\dfrac{1}{b-a}\cdot\dfrac{b^2-a^2}{2}$

$$=\frac{a+b}{2}.$$

$$D(\xi)=\int_{-\infty}^{\infty}x^2f(x)dx-[E(\xi)]^2=\int_a^b\frac{x^2}{b-a}dx-\left(\frac{b+a}{2}\right)^2$$

$$=\frac{1}{b-a}\cdot\frac{b^3-a^3}{3}-\frac{b^2+a^2+2ab}{4}=\frac{b^2+ab+a^2}{3}-\frac{b^2+a^2+2ab}{4}$$

$$=\frac{b^2-2ab+a^2}{12}=\frac{(b-a)^2}{12}.$$

故

$$\left.\begin{aligned}E(\xi)&=\frac{b+a}{2},\\ D(\xi)&=\frac{(b-a)^2}{12}.\end{aligned}\right\}\qquad(3.1.14)$$

例 3.1.6(正态分布) 设 ξ 服从 $N(a,\sigma)$,亦即,密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}\quad(x\in R_1),$$

试求 $E(\xi)$ 与 $D(\xi)$.

解 $E(\xi)=\int_{-\infty}^{\infty}x\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx \quad \left(令\ z=\frac{x-a}{\sigma}\right)$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z+a)e^{-\frac{z^2}{2}}dz$$

$$=\frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=a.$$

$$D(\xi)=\int_{-\infty}^{\infty}(x-a)^2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx \quad \left(令\ z=\frac{x-a}{\sigma}\right)$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2e^{-\frac{z^2}{2}}dz=\frac{\sigma^2}{\sqrt{2\pi}}\left[-ze^{-\frac{z^2}{2}}\Big|_{z=-\infty}^{z=\infty}+\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz\right]$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=\sigma^2.$$

亦即

$$\left.\begin{array}{l}E(\xi)=a,\\ D(\xi)=\sigma^2.\end{array}\right\} \qquad (3.1.15)$$

这说明服从正态分布的随机变数的密度函数由它的数学期望与方差所唯一确定，并且参数 a 与数学期望相同．参数 σ^2 与方差相同．

例 3.1.7(韦布分布)　设随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}\frac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}, & 当\ x>\nu\\ 0, & 当 x\leqslant\nu\end{cases}$$

其中 $m>0,x_0>0$ 及 ν 均为常数，试求 $E(\xi)$ 与 $D(\xi)$.

解 $E(\xi)=\int_{\nu}^{\infty}x\cdot\frac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}dx$

$\left(令\ y=\frac{(x-\nu)^m}{x_0}\right)$

$$=\int_0^{\infty}\left(x_0^{\frac{1}{m}}y^{\frac{1}{m}}+\nu\right)e^{-y}dy=x_0^{\frac{1}{m}}\Gamma\left(\frac{1}{m}+1\right)+\nu.$$

$$D(\xi)=\int_{\nu}^{\infty}x^2\cdot\frac{m}{x_0}(x-\nu)^{m-1}e^{-\frac{(x-\nu)^m}{x_0}}dx-[E(\xi)]^2$$

$$\left(令\ y=\frac{(x-\nu)^m}{x_0}\right)$$

$$=\int_0^{\infty}(x_0^{\frac{1}{m}}y^{\frac{1}{m}}+\nu)^2e^{-y}dy-[E(\xi)]^2$$

$$=\int_0^{\infty}x_0^{\frac{2}{m}}y^{\frac{2}{m}}e^{-y}dy+\nu^2\int_0^{\infty}e^{-y}dy$$

$$+2\nu\int_0^{\infty}x_0^{\frac{1}{m}}y^{\frac{1}{m}}e^{-y}dy-[E(\xi)]^2$$

$$=x_0^{\frac{2}{m}}\Gamma\left(\frac{2}{m}+1\right)+2\nu x_0^{\frac{1}{m}}\Gamma\left(\frac{1}{m}+1\right)+\nu^2$$

$$-x_0^{\frac{2}{m}}\Gamma^2\left(\frac{1}{m}+1\right)-\nu^2-2\nu x_0^{\frac{1}{m}}\Gamma\left(\frac{1}{m}+1\right)$$

$$=x_0^{\frac{2}{m}}\left[\Gamma\left(\frac{2}{m}+1\right)-\Gamma^2\left(\frac{1}{m}+1\right)\right].$$

亦即

$$\left.\begin{aligned}E(\xi)&=x_0^{\frac{1}{m}}\Gamma\left(\frac{1}{m}+1\right)+\nu,\\D(\xi)&=x_0^{\frac{2}{m}}\left[\Gamma\left(\frac{2}{m}+1\right)-\Gamma^2\left(\frac{1}{m}+1\right)\right].\end{aligned}\right\}\tag{3.1.16}$$

例3.1.8(Γ-分布) 设随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}, & x\geq 0;\\0, & x<0,\end{cases}$$

其中 $\alpha>-1$, $\beta>0$ 为常数，试求 $E(\xi)$ 与 $D(\xi)$.

解 $E(\xi)=\int_0^{\infty}x\cdot\frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}dx\qquad\left(令\ y=\frac{x}{\beta}\right)$

$$=\frac{\beta}{\Gamma(\alpha+1)}\int_0^{\infty}y^{\alpha+1}e^{-y}dy=\frac{\beta}{\Gamma(\alpha+1)}\cdot\Gamma(\alpha+2)$$

$$=\beta\cdot(\alpha+1).$$

$$D(\xi)=\int_0^{\infty}x^2\frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}dx-[E(\xi)]^2 \quad \left(令\ y=\frac{x}{\beta}\right)$$

$$=\frac{\beta^2}{\Gamma(\alpha+1)}\int_0^{\infty}y^{\alpha+2}e^{-y}dy-[E(\xi)]^2$$

$$=\frac{\beta^2}{\Gamma(\alpha+1)}\cdot\Gamma(\alpha+3)-\beta^2\cdot(\alpha+1)^2$$

$$=\beta^2\cdot(\alpha+2)\cdot(\alpha+1)-\beta^2(\alpha+1)^2$$

$$=\beta^2(\alpha+1)[(\alpha+2)-(\alpha+1)]=\beta^2(\alpha+1).$$

亦即

$$\left.\begin{aligned}E(\xi)&=\beta(\alpha+1),\\ D(\xi)&=\beta^2(\alpha+1).\end{aligned}\right\} \tag{3.1.17}$$

由(3.1.17)式可见，服从Γ-分布的随机变数的密度函数由它的数学期望与方差所唯一确定.

例 3.1.9(对数正态分布) 设ξ的密度函数为

$$f(x)=\begin{cases}\dfrac{\lg e}{\sqrt{2\pi}\,\sigma x}e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2}, & 当\ x>0;\\ 0, & 当\ x\leqslant 0,\end{cases}$$

其中 $\sigma>0$，μ 为常数，求 $E(\xi)$ 与 $D(\xi)$.

解
$$E(\xi)=\int_0^{\infty}x\cdot\frac{\lg e}{\sqrt{2\pi}\,\sigma x}e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2}dx$$

$$=\frac{10^{\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}10^{\sigma y}e^{-\frac{y^2}{2}}dy$$

$$\left(令\ y=\frac{\lg x-\mu}{\sigma},\ x=10^{\sigma y+\mu}\right)$$

$$=\frac{10^{\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}+\sigma y\ln 10}dy$$

$$=\frac{10^{\mu}}{\sqrt{2\pi}}e^{\frac{\sigma^2\ln^2 10}{2}}\cdot\int_{-\infty}^{\infty}e^{-\frac{(y-\sigma\ln 10)^2}{2}}dy$$

$$=10^{\mu}\cdot e^{\frac{\sigma^2\ln 10}{2}\cdot\ln 10}=10^{\mu+\frac{\sigma^2}{2}\cdot\ln 10}$$

$$= 10^{\mu+\frac{\sigma^2 \ln 10}{2}}.$$

$$D(\xi)=\int_0^{\infty} x^2 \frac{\lg e}{\sqrt{2\pi}\,\sigma x} e^{-\frac{1}{2}\left(\frac{\lg x-\mu}{\sigma}\right)^2} dx-[E(\xi)]^2$$

$$\left(令\ y=\frac{\lg x-\mu}{\sigma}\right)$$

$$=\frac{10^{2\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} 10^{2\sigma y} e^{-\frac{y^2}{2}} dy-[E(\xi)]^2$$

$$=10^{2\mu+2\sigma^2\cdot\frac{1}{\lg e}}-\left(10^{\mu+\frac{\sigma^2}{2}\cdot\frac{1}{\lg e}}\right)^2$$

$$=10^{2\mu+\sigma^2\cdot\frac{1}{\lg e}}\cdot\left[10^{\sigma^2\frac{1}{\lg e}}-1\right]$$

$$=10^{2\mu+\sigma^2\ln 10}\left(10^{\sigma^2\ln 10}-1\right).$$

即

$$\left.\begin{aligned} E(\xi)&=10^{\mu+\frac{\sigma^2}{2}\ln 10},\\ D(\xi)&=10^{2\mu+\sigma^2\ln 10}\left(10^{\sigma^2\ln 10}-1\right).\end{aligned}\right\} \qquad (3.1.18)$$

故服从对数正态分布的随机变数的密度函数，也由它的数学期望与方差所唯一确定.

例 3.1.10(χ^2-分布)　设随机变数 χ^2 的密度函数为

$$f_{\chi^2}(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\geqslant 0;\\ 0, & 其它,\end{cases}$$

其中 n 为正整数，试求 $E(\xi)$ 与 $D(\xi)$.

解　$$E(\xi)=\int_0^{\infty}\frac{x}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx \quad \left(令\ y=\frac{x}{2}\right)$$

$$=\frac{2}{\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty}y^{\frac{n}{2}}e^{-y}dy=\frac{2}{\Gamma\left(\frac{n}{2}\right)}\cdot\Gamma\left(\frac{n}{2}+1\right)=n;$$

$$D(\xi)=\int_0^{\infty}\frac{x^2}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx-[E(\xi)]^2\quad\left(令\ y=\frac{x}{2}\right)$$

$$=\frac{4}{\Gamma\left(\frac{n}{2}\right)}\int_0^{\infty}y^{\frac{n}{2}+1}e^{-y}dy-[E(\xi)]^2$$

$$=\frac{4}{\Gamma\left(\frac{n}{2}\right)}\cdot\Gamma\left(\frac{n}{2}+2\right)-n^2=4\left(\frac{n}{2}+1\right)\cdot\frac{n}{2}-n^2=2n.$$

即

$$\left.\begin{aligned}E(\xi)&=n,\\D(\xi)&=2n.\end{aligned}\right\}\tag{3.1.19}$$

例 3.1.11（t-分布） 设随机变数 ξ 之密度函数为

$$f_\xi(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},$$

其中 n 为正整数，试求 $E(\xi)$ 与 $D(\xi)$。

解

$$E(\xi)=\int_{-\infty}^{\infty}\frac{x\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx=0\quad(n>1)$$

（∵ 被积函数是奇函数）；

$$D(\xi)=\int_{-\infty}^{\infty}\frac{x^2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx$$

$$=2\int_0^{\infty}\frac{x^2\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx$$

$$=2n\int_0^{\infty}\frac{\frac{x^2}{n}\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx$$

$$=2n\left[\int_0^{\infty}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n-1}{2}}dx\right.$$

$$\left.-\int_0^{\infty}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\cdot\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}dx\right]$$

$$=\frac{2n\frac{n-1}{2}}{\frac{n-2}{2}}\int_0^{\infty}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{(n-2)\pi}\,\Gamma\left(\frac{n-2}{2}\right)}\cdot$$

$$\left(1+\frac{\frac{n-2}{n}x^2}{n-2}\right)^{-\frac{(n-2)+1}{2}}d\left(\sqrt{\frac{n-2}{n}}x\right)-n$$

$$=n\cdot\frac{n-1}{n-2}-n=n\left(\frac{n-1}{n-2}-1\right)=\frac{n}{n-2},\quad (当\ n>2\ 时).$$

$$\left(这里用到了\int_{-\infty}^{\infty}f_t(x)dx=1.\right)$$

即得

$$\left.\begin{aligned}&E(\xi)=0 && (当\ n>1\ 时),\\&D(\xi)=\frac{n}{n-2} && (当\ n>2\ 时).\end{aligned}\right\}\qquad(3.1.20)$$

当 $n=1$ 时, t-分布就是 $\lambda=1$, $\mu=0$ 时的哥西分布. 期望、方差不存在(见例 3.1.13).

当 $n=2$ 时,应另行计算 $D(\xi)$(读者自行计算得 $D(\xi)=\infty$,故 $D(\xi)$ 不存在).

例 3.1.12(F-分布) 设随机变数 ξ 的密度函数为

$$
f_{\xi}(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & \text{当 } x>0;\\ 0, & \text{当 } x\leqslant 0,\end{cases}
$$

其中 m、n 为正整数，试求 $E(\xi)$ 与 $D(\xi)$。

解

$$
E(\xi)=\int_0^{\infty}\frac{x\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}dx
$$

$$
=\int_0^{\infty}\frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}m^{\frac{m}{2}}n^{\frac{n}{2}}\frac{x^{\frac{m}{2}}}{(mx+n)^{\frac{m+n}{2}}}dx\quad\left(\text{令 } z=\frac{m}{n}x\right)
$$

$$
=\int_0^{\infty}\frac{\Gamma\left(\dfrac{m-n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{2}{n}\right)}\frac{n}{m}\frac{z^{\frac{m}{2}}}{(1+z)^{\frac{m+n}{2}}}dz\quad\left(\text{令}\frac{1}{1+z}=y\right)
$$

$$
=\frac{n}{m}\frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\int_0^1 y^{\frac{n}{2}-2}(1-y)^{\frac{m}{2}}dy
$$

$$
=\frac{n}{m}\frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}B\left(\frac{n}{2}-1,\frac{m}{2}+1\right)^{①}\quad(\text{当 } n>2 \text{ 时})
$$

$$
=\frac{n}{m}\frac{\Gamma\left(\dfrac{n}{2}-1\right)\Gamma\left(\dfrac{m}{2}+1\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}=\frac{n}{m}\frac{\dfrac{m}{2}}{\dfrac{n}{2}-1}
$$

---

① $B(m,n)=\int_0^1 x^{m-1}(1-x)^{n-1}dx=\dfrac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}$，这是数学分析中熟知的公式 $(m>0,n>0)$。

$$=\frac{n}{n-2}.\quad (当\ n>2\ 时)$$

同理可得

$$D(\xi)=\frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}\quad (当\ n>4\ 时).$$

即

$$\left.\begin{aligned}E(\xi)&=\frac{n}{n-2}\quad (n>2),\\ D(\xi)&=\frac{2n^2(n+m-2)}{m(n-2)^2(n-4)}\quad (n>4).\end{aligned}\right\}\qquad (3.1.21)$$

值得注意的是，并非所有的随机变数 ξ，都有数学期望和方差存在.

例 3.1.13 （数学期望和方差不存在的例子——哥西分布）

设 ξ 的密度函数为

$$f(x)=\frac{1}{\pi}\frac{1}{(1+x^2)}\quad (x\in R_1)$$

则 $E(\xi)$ 不存在，因而 $D(\xi)$ 也不存在.

解 由于 $\int_{-A}^{A}|x|\frac{1}{\pi}\frac{1}{(1+x^2)}dx=\frac{2}{2\pi}\int_0^A\frac{d(1+x^2)}{1+x^2}=$ $\frac{1}{\pi}\ln(1+x^2)\Big|_{x=0}^{x=A}\to\infty,\ (A\to\infty)$，

故 $E(\xi)$ 不存在，因而 $D(\xi)$ 也不存在.

二、一般的随机变数的数学期望与方差的定义和性质

我们在前一段，对于离散型和连续型随机变数给出了数学期望和方差的定义，给出了定义 3.1.1 及定义 3.1.2. 在那里给出了两种不同的定义形式. 但对于一般的随机变数（既非离散型也非连续型的）数学期望与方差等没有给出它的定义. 因此需要给出更一般的定义来解决上述提出的问题. 为此，我们可以藉助于黎曼-斯梯阶积分①（简称 R-S 积分）来定义随机变数的数学期望

① 见附录 III.

和方差.

定义 3.1.3 （数学期望） 设 ξ 为随机变数，其分布函数为 $F(x)$，若

$$\int_{-\infty}^{\infty}|x|dF(x)<\infty,$$

则记

$$E(\xi)=\int_{-\infty}^{\infty}xdF(x). \qquad (3.1.22)$$

并称 $E(\xi)$ 为 ξ 的数学期望.

(3.1.22)式积分为 R-S 积分，由于 $F(x)$ 是 ξ 的分布函数，故单调上升且有界，因而(3.1.22)式是有意义的.

学过实变函数的读者会知道，黎曼积分 $\int_a^b f(x)dx$ 的可积函数类较小，勒贝格积分则大大地拓广了其可积函数类. 关于 R-S 积分也有类似情形. 若引入 L-S 积分[①]，则数学期望存在的随机变数 ξ 便会多许多.

我们也可以不依赖于 L-S 积分，直接作出 ξ 的数学期望的定义（虽然这一定义实质上也可以化为 L-S 积分的情形）. 先设 ξ 为有界的随机变数，亦即存在常数 k，使得

$$|\xi(\omega)|\leqslant k, \qquad \text{对一切 } \omega\in\Omega.$$

这时，我们可以令

$$E\xi=\lim_{n\to\infty}\sum_{k=-\infty}^{\infty}\frac{k}{2^n}P\left(\frac{k}{2^n}<\xi\leqslant\frac{k+1}{2^n}\right).$$

应注意，由于 ξ 之有界性，上述级数对每个 n，实际是有限和. 并且可以证明，此时 $E\xi$ 与 $E|\xi|$ 皆存在. 在一般情形，令

$$\xi_N(\omega)=\xi(\omega)\cdot I_{(|\xi|\leqslant N)}(\omega),$$

其中

$$I_{(|\xi|\leqslant N)}(\omega)=\begin{cases}1, & \text{当 } |\xi(\omega)|\leqslant N,\\ 0, & \text{当 } |\xi(\omega)|>N.\end{cases}$$

---

① 见《测度与概率基础》（中山大学《测度与概率基础》编写组）第四章 §5.

若$\lim\limits_{N\to\infty}E|\xi_N|$存在且$<+\infty$，则令

$$E\xi=\lim_{N\to\infty}E\xi_N.$$

这就得到了一般随机变数ξ的数学期望$E\xi$的定义．由此定义出发，可以得到书中一切有关数学期望的结果①．但这些论证都已超出了本书范围，兹从略

当ξ为离散型随机变数时，则(3.1.22)式变为

$$E(\xi)=\sum_i x_i P\{\xi=x_i\}. \tag{3.1.23}$$

事实上，当$F(x)$取值c_k，($c_{k-1}<c_k$)$k=1,2,\cdots$时，由附录III中的积分定理2.3立即可得．由（3.1.23）式看到，此时数学期望$E(\xi)$与定义3.1.1一致．

当ξ为连续型随机变数时，则(3.1.22)式变为

$$E(\xi)=\int_{-\infty}^{\infty}xf(x)dx. \tag{3.1.24}$$

事实上，由附录III中的积分定理2.2可得．由(3.1.24)可以看到，此时$E(\xi)$与定义3.1.2一致．

为了引入随机变数ξ的方差，以及讨论数学期望与方差的性质，我们引入下列重要而有用的定理．

定理3.1.1　设$F(x)$为随机变数ξ的分布函数，$g(x)$为R_1上的连续函数．若$\int_{-\infty}^{\infty}|g(x)|dF(x)<\infty$，则$g(\xi)$的数学期望$E[g(\xi)]$存在②，且

$$E[g(\xi)]=\int_{-\infty}^{\infty}g(x)dF(x). \tag{3.1.25}$$

证　只对$g(x)$是x的严格上升函数这一情形加以证明．令$\eta=g(\xi)$．则η的分布函数为

---

① M. Loeve, Probability Theory, 2nd ed. D. Van Nostrard Co. N. Y. (1960).

② 由定理2.3.2知，$g(\xi)$必为随机变数。

$$F_\eta(y) \quad P(g(\xi)<y)=P(\xi<g^{-1}(y))$$
$$=F(g^{-1}(y)).$$

由定义 3.1.3 得

$$Eg(\xi)=E\eta=\int_{-\infty}^{\infty} y dF_\eta(y)=\int_{-\infty}^{\infty} y dF(g^{-1}(y))$$
$$=\int_{-\infty}^{\infty} g(x)dF(x). \quad (令\ y=g(x)). ▌$$

若令 $\eta=g(\xi)$，则(3.1.25) 的左端 $E[g(\xi)]=E(\eta)$. 按定义 3.1.3, $E(\eta)=\int_{-\infty}^{\infty} ydF_\eta(y)$，其中 $F_\eta(y)$ 为 η 的分布函数. 因此，求 $E(\eta)$ 必须先求得 $F_\eta(y)$. 定理 3.1.1 告诉我们，在定理的条件下，可以直接通过 ξ 的分布函数 $F(x)$ 按(3.1.25)式右边的积分来得. 这对我们计算随机变数的函数的数字特征十分方便.

性质 3.1.1 设随机变数 ξ 有数学期望 $E(\xi)$，则 $\eta=a\xi+b$，(a，b 均为常数)的数学期望是

$$E(\eta)=aE(\xi)+b,$$

特别当 $a=0$ 时有 $E(b)=b$. 即常数 b 的数学期望就是它自己本身.

证 设 ξ 的分布函数为 $F(x)$，按定理 3.1.1 及 R-S 积分的性质得

$$E(\eta)=E[a\xi+b]=\int_{-\infty}^{\infty}(ax+b)dF(x)$$
$$=a\int_{-\infty}^{\infty} xdF(x)+b\int_{-\infty}^{\infty} dF(x)$$
$$=aE(\xi)+b[F(\infty)-F(-\infty)]$$
$$=aE(\xi)+b.$$

从上式，特别当 $a=0$，则有 $E(\eta)=E(b)=b$. ▌

定义 3.1.4 设随机变数 ξ 的分布函数为 $F(x)$，数学期望为 $E(\xi)$，若 $E[\xi-E(\xi)]^2<\infty$，记 $D(\xi)=E[\xi-E(\xi)]^2$，称 $D(\xi)$ 为 ξ 的方差.

由定理 3.1.1 有

$$D(\xi)=E[\xi-E(\xi)]^2=\int_{-\infty}^{\infty}[x-E(\xi)]^2dF(x). \quad (3.1.26)$$

由定义 3.1.4 知，方差 $D(\xi)$ 存在蕴含了数学期望存在，且方差 $D(\xi)\geqslant 0$. 与数学期望情形类似，当 ξ 为离散型随机变数时，(3.1.26) 式为

$$D(\xi)=E[\xi-E(\xi)]^2=\sum_{i=0}^{\infty}[x_i-E(\xi)]^2p_i. \quad (3.1.27)$$

当 ξ 为连续型随机变数时 (3.1.26) 式变为

$$D(\xi)=E[\xi-E(\xi)]^2=\int_{-\infty}^{\infty}[x-E(\xi)]^2f(x)\,dx. \quad (3.1.28)$$

性质 3.1.2 设 ξ 为一随机变数，$E(\xi^2)<\infty$. 则 $E(\xi)$ 及 $D(\xi)$ 存在且

$$D(\xi)=E(\xi^2)-[E(\xi)]^2; \quad (3.1.29)$$

$$\left[\int_{-\infty}^{\infty}x\,dF(x)\right]^2\leqslant\int_{-\infty}^{\infty}x^2dF(x). \quad (3.1.30)$$

((3.1.30) 式是许瓦兹不等式的一个特殊情形).

证 由定理 3.1.1 及 R-S 积分性质，并利用熟知的不等式 $|x|\leqslant 1+|x|^2$ 有

$$E|\xi|=\int_{-\infty}^{\infty}|x|dF(x)\leqslant\int_{-\infty}^{\infty}[1+|x|^2]dF(x)$$
$$=\int_{-\infty}^{\infty}dF(x)+\int_{-\infty}^{\infty}x^2dF(x)=1+E(\xi^2)<\infty.$$

故 $E(\xi)$ 存在. 另一方面：

$$D(\xi)=E[\xi-E(\xi)]^2=\int_{-\infty}^{\infty}[x-E(\xi)]^2dF(x)$$
$$=\int_{-\infty}^{\infty}x^2dF(x)-2E(\xi)\int_{-\infty}^{\infty}x\,dF(x)$$
$$+[E(\xi)]^2\int_{-\infty}^{\infty}dF(x)$$

$$=E(\xi^2)-[E(\xi)]^2<\infty.$$

最后由 $D(\xi)\geqslant 0$ 即得(3.1.30) ▮

性质 3.1.3 设随机变数 ξ 的分布函数为 $F(x)$，方差 $D(\xi)$ 存在，则 $\eta=a\xi+b$ 的方差

$$D(\eta)=D(a\xi+b)=a^2D(\xi). \tag{3.1.31}$$

特别当 $a=0$，则有 $D(b)=0$.

证 由(3.1.26)式和性质 3.1.1 得

$$\begin{aligned}D(\eta)=D(a\xi+b)&=E[a\xi+b-E(a\xi+b)]^2\\&=E[a\xi+b-aE(\xi)-b]^2\\&=E\{a[\xi-E(\xi)]\}^2\\&=\int_{-\infty}^{\infty}a^2[x-E(\xi)]^2dF(x)\\&=a^2\int_{-\infty}^{\infty}[x-E(\xi)]^2dF(x)=a^2D(\xi).\end{aligned}$$ ▮

在理论研究和实际工作中，为了方便计算或简化证明，往往对随机变数进行所谓"标准化"：当随机变数 ξ 的数学期望 $E(\xi)$ 和方差 $D(\xi)$ 都存在时，考虑它的标准化随机变数：

$$\xi^*=\frac{\xi-E(\xi)}{\sqrt{D(\xi)}},$$

由性质 3.1.1 及性质 3.1.3，得

$$E(\xi^*)=\frac{1}{\sqrt{D(\xi)}}[E(\xi)-E(\xi)]=0,$$

$$D(\xi^*)=\frac{1}{D(\xi)}D(\xi)=1.$$

例如，随机变数 ξ 服从正态 $N(a,\ \sigma)$，由例 3.1.6 知 $E(\xi)=a$，$D(\xi)=\sigma^2$. 则它的标准化随机变数 $\xi^*=\frac{\xi-a}{\sigma}$ 的数学期望为 $E(\xi^*)=0$，方差 $D(\xi^*)=1$，即 ξ^* 服从标准正态 $N(0,1)$.

性质 3.1.4 函数 $f(x)=E[(\xi-x)^2]$，$x\in R_1$，当 $x=E(\xi)$ 时

达到最小值.

证 ∵ $f(x)=E[(\xi-x)^2]=E(\xi^2)-2xE(\xi)+x^2$

∴ $\frac{df(x)}{dx}=2x-2E(\xi)$.

故 $x=E(\xi)$ 为 $f(x)$ 的一个逗留点,再由

$$\frac{d^2f(x)}{dx^2}=2$$

知 $f(x)$ 为凹函数,故 $f(x)$ 在点 $x=E(\xi)$ 达到最小值,并且其最小值为

$$f[E(\xi)]=E[\xi-E(\xi)]^2. \blacksquare$$

性质 3.1.4 说明:随机变数对于其数学期望的偏离程度比它关于其他任何值的偏离程度都来得小. 假如我们用转动惯量来解释方差,则这一物理直观就更清楚了. 因为,质量分布于横坐标轴的物体相对于质量重心的纵轴的转动惯量总是比相对于其他点的纵轴的转动惯量小.

定理 3.1.2 (车贝谢夫不等式) 若随机变数 ξ 的方差 $D(\xi)$ 存在,则对任意 $\varepsilon>0$,有

$$P\{|\xi-E(\xi)|\geqslant\varepsilon\}\leqslant\frac{D(\xi)}{\varepsilon^2}. \tag{3.1.32}$$

证 设 ξ 的分布函数为 $F(x)$,则

$$\begin{aligned}P\{|\xi-E(\xi)|\geqslant\varepsilon\}&=\int_{|x-E(\xi)|\geqslant\varepsilon}dF(x)\\&\leqslant\int_{|x-E(\xi)|\geqslant\varepsilon}\frac{[x-E(\xi)]^2}{\varepsilon^2}dF(x)\\&\leqslant\frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}[x-E(\xi)]^2dF(x)\\&=\frac{D(\xi)}{\varepsilon^2}. \blacksquare\end{aligned}$$

定理 3.1.2 说明,若 ξ 的方差小,则事件 $\{|\xi-E(\xi)|\geqslant\varepsilon\}$ 发

生的概率就小. 换句话说，事件$\{|\xi-E(\xi)|<\varepsilon\}$ 发生的概率就大，也就是说随机变数ξ取的值基本上集中于$E(\xi)$附近. 这进一步给我们说明了方差的意义. 同时，当$D(\xi)$已知时，(3.1.32)式给出一个随机变数ξ与$E(\xi)$的偏差不小于ε的概率的估计式.

推论 若$D(\xi)=0$, 则ξ以概率为1地等于它的数学期望$E(\xi)$，即$P\{\xi=E(\xi)\}=1$.

证 由假设$D(\xi)=0$，故对任意$n=1,2,\cdots$，有

$$P\left\{|\xi-E(\xi)|\geqslant\frac{1}{n}\right\}\leqslant\frac{D(\xi)}{\left(\frac{1}{n}\right)^2}=0.$$

又因为

$$\{|\xi-E(\xi)|\neq 0\}=\bigcup_{n=1}^{\infty}\left\{|\xi-E(\xi)|\geqslant\frac{1}{n}\right\}.$$

故

$$P\{|\xi-E(\xi)|\neq 0\}\leqslant\sum_{n=1}^{\infty}P\left\{|\xi-E(\xi)|\geqslant\frac{1}{n}\right\}=0.$$

即得

$$P\{\xi=E(\xi)\}=1-P\{|\xi-E(\xi)|\neq 0\}=1.\blacksquare$$

这推论说明，当方差为0时，随机变数ξ(以概率为1地)取值集中于数学期望这一值上，故方差刻划了随机变数ξ围绕它的数学期望的偏差程度.

§3.2 矩

随机变数的数字特征除了上述数学期望与方差外，还有其他数字特征. 这一节我们介绍随机变数的各阶矩（原点矩、中心矩等)，它们在数理统计中有重要的应用.

定义 3.2.1 设ξ为随机变数. 若$E(|\xi|^k)<\infty$

记$\nu_k=E(\xi^k)$，$\alpha_k=E(|\xi|^k)$

则称 ν_k 为 ξ 的 k 阶原点矩，并称 α_k 为 ξ 的 k 阶原点绝对矩.

又若 $E(\xi)$ 存在,且 $E(|\xi-E(\xi)|^k)<\infty$,记

$$\mu_k=E[\xi-E(\xi)]^k.$$

则称 μ_k 为 ξ 的 k 阶中心矩，并称 $\beta_k=E|\xi-E(\xi)|^k$ 为 ξ 的 k 阶中心绝对矩.

显然有 $\nu_0=1,\nu_1=E(\xi),\mu_0=1,\mu_1=0,\mu_2=D(\xi)$.

定理 3.2.1 （马尔可夫不等式）若随机变数 ξ 的 k 阶原点绝对矩 α_k 存在,则对任意 $\varepsilon>0$ 有

$$P\{|\xi|\geqslant\varepsilon\}\leqslant\frac{\alpha_k}{\varepsilon^k}. \tag{3.2.1}$$

证 设 ξ 的分布函数为 $F(x)$,

$$P\{|\xi|\geqslant\varepsilon\}=\int_{|x|\geqslant\varepsilon}dF(x)\leqslant\int_{|x|\geqslant\varepsilon}\frac{|x|^k}{\varepsilon^k}dF(x)$$

$$\leqslant\frac{1}{\varepsilon^k}\int_{-\infty}^{\infty}|x|^k dF(x)=\frac{1}{\varepsilon^k}\alpha_k. \blacksquare$$

上述定义的原点矩与中心矩之间有如下关系:

$$\mu_n=E[\xi-E(\xi)]^n=\sum_{k=0}^{n}C_n^k(-1)^{n-k}E(\xi^k)[E(\xi)]^{n-k}$$

$$=\sum_{k=0}^{n}(-1)^{n-k}C_n^k(\nu_1)^{n-k}\nu_k. \tag{3.2.2}$$

故可得

$$\left.\begin{aligned}&\mu_0=1,\\&\mu_1=0,\\&\mu_2=\nu_2-\nu_1^2,\\&\mu_3=\nu_3-3\nu_2\nu_1+2\nu_1^3,\\&\mu_4=\nu_4-4\nu_3\nu_1+6\nu_2\nu_1^2-3\nu_1^4,\\&\mu_5=\nu_5-5\nu_4\nu_1+10\nu_3\nu_1^2-10\nu_2\nu_1^3+4\nu_1^5,\\&\cdots\cdots\cdots\cdots\end{aligned}\right\} \tag{3.2.3}$$

上式说明，由原点矩可求中心矩．反过来，由中心矩也可求原点矩．事实上（见§3.4）

$$\nu_n = E(\xi^n) = E[(\xi-\nu_1)+\nu_1]^n = \sum_{k=0}^{n} C_n^k \nu_1^k E[(\xi-\nu_1)^{n-k}]$$

$$= \sum_{k=0}^{n} C_n^k \nu_1^k \mu_{n-k}. \tag{3.2.4}$$

故可得

$$\left.\begin{aligned} &\nu_0 = 1,\\ &\nu_1 = E(\xi),\\ &\nu_2 = \mu_2 + (E(\xi))^2,\\ &\nu_3 = \mu_3 + 3\nu_1\mu_2 + \nu_1^3,\\ &\nu_4 = \mu_4 + 4\nu_1\mu_3 + 6\nu_1^2\mu_2 + \nu_1^4,\\ &\nu_5 = \mu_5 + 5\nu_1\mu_4 + 10\nu_1^2\mu_3 + 10\nu_1^3\mu_2 + \nu_1^5,\\ &\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{3.2.5}$$

例 3.2.1 设随机变数 ξ 服从 $N(a,\sigma)$，则 ξ 的 k 阶中心矩为

$$\mu_k = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)^k e^{-\frac{(x-a)^2}{2\sigma^2}}dx = \frac{\sigma^k}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^k e^{-\frac{y^2}{2}}dy.$$

当 k 为奇数时，由于被积函数为奇函数，故 $\mu_k=0$.

当 k 为偶数时，令 $y^2=2z$，则

$$\mu_k = \frac{1}{\sqrt{2\pi}}2\sigma^k\int_0^{\infty} y^k e^{-\frac{y^2}{2}}dy = \sqrt{\frac{2}{\pi}}\sigma^k 2^{\frac{k-1}{2}}\int_0^{\infty} z^{\frac{k-1}{2}}e^{-z}dz$$

$$= \frac{1}{\sqrt{\pi}}\sigma^k 2^{\frac{k}{2}}\Gamma\left(\frac{k+1}{2}\right)$$

$$= \sigma^k(k-1)(k-3)\cdots 1 \quad (\text{当 } k\geqslant 2). \tag{3.2.6}$$

因而有

$$\mu_0=1,\ \mu_2=\sigma^2,\ \mu_4=3\sigma^4,\ \mu_6=15\sigma^6,\ \cdots$$

下面我们求 k 阶中心绝对矩 β_k.

显然，当 k 为偶数 $\beta_k=\mu_k$.

当 k 为奇数时，与求 μ_k 时类似，有

$$\beta_k=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}|x-a|^k e^{-\frac{(x-a)^2}{2\sigma^2}}dx$$

$$=\sqrt{\frac{2}{\pi}}\sigma^k\int_0^{\infty}y^k e^{-\frac{y^2}{2}}dy=\frac{1}{\sqrt{\pi}}2^{\frac{k}{2}}\left(\frac{k-1}{2}\right)!\,\sigma^k. \qquad (3.2.7)$$

因而有

$$\beta_1=\sqrt{\frac{2}{\pi}}\sigma,\ \beta_3=\frac{2\sqrt{2}}{\sqrt{\pi}}\sigma^3,\ \beta_5=\frac{8\sqrt{2}}{\sqrt{\pi}}\sigma^5,\cdots$$

§3.3 多维随机变数的数字特征

与一维随机变数类似，多维随机变数的数学期望与方差的定义如下：

定义 3.3.1 设 n 维随机变数 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的分布函数为 $F(x_1,x_2,\cdots,x_n)$．ξ_k 的分布函数为 $F_k(x_k)$，$(k=1,2,\cdots,n)$ (ξ_i,ξ_k) 的联合分布函数为 $F_{ik}(x_i,x_k)$．

设所有下面出现的积分都绝对收敛．

记

$$a_k=\int_{-\infty}^{\infty}x_k dF_k(x_k),\ (k=1,2,\cdots,n). \qquad (3.3.1)$$

称 $(a_1,a_2,\cdots,a_n)$ 为 n 维随机变数 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的数学期望．

记 $b_{ik}=E\{[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\}$

$$=\begin{cases}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x_i-E(\xi_i)][x_k-E(\xi_k)]dF_{ik}(x_i,x_k) & \text{当 } i\neq k,\\ \int_{-\infty}^{\infty}[x_i-E(\xi_i)]^2dF_i(x_i) & \text{当 } i=k\end{cases}$$

$$(i,k=1,2,\cdots,n), \qquad (3.3.2)$$

称 $b_{ik}(k\neq i)$ 为随机变数 ξ_i 与 ξ_k 的二阶混合中心矩，统称 b_{ik}（不论 $i=k$ 与否）为协方差．（有时又写作 $b_{ik}=\text{cov}(\xi_i,\xi_k)$．）

称方阵

$$\Sigma=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \tag{3.3.3}$$

为 n 维随机变数$(\xi_1,\xi_2,\cdots,\xi_n)$的协方差矩阵.

协方差矩阵中的元素 b_{ik} 有如下性质.

性质 3.3.1 1) $b_{kk}=D(\xi_k)\quad k=1,2,\cdots,n$;

2) $b_{ik}=b_{ki}\quad i,k=1,2,\cdots,n$;

这表明协方差矩阵是对称矩阵.

3) $b_{ik}^2\leqslant b_{ii}\cdot b_{kk}$

$$i,k=1,2,\cdots,n.$$

证 1) $b_{kk}=\int_{-\infty}^{\infty}[x_k-E(\xi_k)]^2dF_k(x_k)=D(\xi_k)$;

$$k=1,2,\cdots,n.$$

2) 由定义 3.3.1, $b_{ik}=b_{ki}$ 是显然的;

3) 由 R-S 积分的许瓦兹不等式即得. ∎

性质 3.3.2 对任意实数 t_i, $(i=1,2,\cdots,n)$有

$$\sum_{i=1}^{n}\sum_{k=1}^{n}b_{ik}t_it_k\geqslant 0.\ \text{即}\ T'\Sigma T\geqslant 0;$$

其中 Σ 定义如(3.3.3), $T=\begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}$, T' 是 T 的转置, 亦即 $T'=(t_1,\cdots,t_n)$.

证 $\quad\sum_{i=1}^{n}\sum_{k=1}^{n}b_{ik}t_it_k$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}t_it_k[x_i-E(\xi_i)][x_k-E(\xi_k)]dF_{ik}(x_i,x_k)$$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}t_i t_k[x_i-E(\xi_i)][x_k-E(\xi_k)]$$

$$\cdot dF(x_1,x_2,\cdots,x_n)$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\sum_{i=1}^{n}t_i[x_i-E(\xi_i)]\right\}^2 dF(x_1,x_2,\cdots,x_n)$$

$$\geqslant 0. \blacksquare$$

性质 3.3.2 说明(3.3.3) 式的协方差矩阵是非负定的. 从矩阵论的二次型理论和性质 3.3.2 可知, 对任意 $k=1, 2,\cdots,n$, 行列式

$$\begin{vmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{vmatrix} \geqslant 0. \qquad (3.3.4)$$

特别取 $k=2$ 得:

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = b_{11}b_{22}-b_{12}b_{21}\geqslant 0.$$

注意到

$$b_{12}=b_{21}.$$

即得

$$b_{12}^2\leqslant b_{11}b_{22}.$$

这正是性质 3.3.1 之 3) 的特殊情况.

同§3.1 讨论一样, 当 $\xi_1,\xi_2,\cdots,\xi_n$ 均是离散型随机变数时, 设其分布列为:

ξ_i	$x_1^{(i)}$	$x_2^{(i)}$	$\cdots$	$x_l^{(i)}$	$\cdots$
$P\{\xi_i=x_l^{(i)}\}$	$p_1^{(i)}$	$p_2^{(i)}$	$\cdots$	$p_l^{(i)}$	$\cdots$

其中对任意 $i=1,2,\cdots,n$, $p_l^{(i)}\geqslant 0$; $\sum_l p_l^{(i)}=1$.

记 $P(x_l^{(i)}, x_m^{(k)}) = P\{\xi_i = x_l^{(i)}, \xi_k = x_m^{(k)}\}$

$$i, k = 1, 2, \cdots, n, \ l, m = 1, 2, \cdots,$$

则(3.3.1)及(3.3.2)变为

$$a_i = E(\xi_i) = \sum_l x_l^{(i)} p_l^{(i)}, \ i = 1, 2, \cdots, n. \tag{3.3.5}$$

$$b_{ik} = \begin{cases} \sum_m \sum_l [x_l^{(i)} - E(\xi_i)][x_m^{(k)} - E(\xi_k)] P(x_l^{(i)}, x_m^{(k)}), & i \neq k \\ \sum_l [x_l^{(i)} - E(\xi_i)]^2 p_l^{(i)}, \ i = k, \end{cases}$$

$$i, k = 1, 2, \cdots, n. \tag{3.3.6}$$

当 $\xi_1, \xi_2, \cdots, \xi_n$ 均为连续型随机变数时，设其 n 维密度函数为 $f(x_1, x_2, \cdots, x_n)$，其一维及二维密度函数为 $f_i(x_i)$ 及 $f_{ik}(x_i, x_k)$. 则(3.3.1)及(3.3.2)式变为

$$a_i = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i = E(\xi_i), \ i = 1, 2, \cdots, n. \tag{3.3.7}$$

$$b_{ik} = \begin{cases} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_i - E(\xi_i)][x_k - E(\xi_k)] f_{ik}(x_i, x_k) dx_i dx_k, \ i \neq k, \\ \int_{-\infty}^{\infty} [x_i - E(\xi_i)]^2 f_i(x_i) dx_i, \ 当 i = k, \end{cases}$$

$$i, k = 1, 2, \cdots, n. \tag{3.3.8}$$

例 3.3.1 （二维二点分布） 设 ξ, η 的分布列由下表给出：

ξ \ η	0	1
0	$1-p$	0
1	0	p

则

$$E(\xi) = 0 \cdot [(1-p) + 0] + 1 \cdot (0 + p) = p,$$
$$E(\eta) = 0 \cdot [(1-p) + 0] + 1 \cdot (0 + p) = p;$$

故(ξ, η)的数学期望是(p, p)（其中 $1 \geq p \geq 0$）.

$$b_{12} = (0-p) \cdot (0-p)(1-p) + (0-p)(1-p) \cdot 0$$

$$+(1-p)(0-p)\cdot 0+(1-p)(1-p)\cdot p$$
$$=p^2(1-p)+(1-p)^2p=p(1-p)(p+1-p)$$
$$=p(1-p),$$
$$b_{11}=D(\xi_1)=(0-p)^2(1-p)+(1-p)^2p=p(1-p),$$
$$b_{22}=D(\xi_2)=p(1-p),\ b_{21}=b_{12},$$

故协方差矩阵为

$$\begin{pmatrix} p(1-p) & p(1-p) \\ p(1-p) & p(1-p) \end{pmatrix}.$$

例 3.3.2（均匀分布） 设(ξ,η)的密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{(b_1-a_1)(b_2-a_2)}, & \text{当 } a_1\leqslant x\leqslant b_1 \quad a_2\leqslant y\leqslant b_2; \\ 0, & \text{其它}\end{cases}$$

求数学期望和协方差矩阵。

解 由例 3.1.5 得

$$a_1=E(\xi)=\int_{a_1}^{b_1}x\cdot\frac{1}{b_1-a_1}dx=\frac{b_1+a_1}{2},$$

$$a_2=E(\eta)=\frac{b_2+a_2}{2},$$

$$b_{11}=D(\xi)=\frac{(b_1-a_1)^2}{12},\ b_{22}=D(\eta)=\frac{(b_2-a_2)^2}{12},$$

$$b_{12}=b_{21}=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\left(x-\frac{b_1+a_1}{2}\right)\cdot\left(y-\frac{b_2+a_2}{2}\right)\cdot\frac{dydx}{(b_1-a_1)(b_2-a_2)}$$
$$=\frac{1}{(b_1-a_1)(b_2-a_2)}\int_{a_1}^{b_1}\left(x-\frac{b_1+a_1}{2}\right)dx\int_{a_2}^{b_2}\left(y-\frac{b_2+a_2}{2}\right)dy=0.$$

故(ξ,η)的数学期望和协方差矩阵分别为

$$\left.\begin{aligned}&(E(\xi),E(\eta))=\left(\frac{b_1+a_1}{2},\frac{b_2+a_2}{2}\right),\\ &\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}\end{pmatrix}=\begin{pmatrix}\dfrac{(b_1-a_1)^2}{12} & 0 \\ 0 & \dfrac{(b_2-a_2)^2}{12}\end{pmatrix}.\end{aligned}\right\} \tag{3.3.9}$$

例 3.3.3(正态分布) 设(ξ,η)的密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left[-\frac{1}{2(1-r^2)}\left\{\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\right\}\right],$$

其中$\sigma_1,\sigma_2>0,m_1,m_2$均为常数，$-1<r<1$. 求数学期望与协方差矩阵.

解 由例 3.1.6 的结果知$E(\xi)=m_1,E(\eta)=m_2$,

$$b_{11}=D(\xi)=\sigma_1^2,\qquad b_{22}=D(\eta)=\sigma_2^2,$$

$$b_{12}=b_{21}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_1)(y-m_2)\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\cdot\exp\left\{-\frac{\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}}{2(1-r^2)}\right\}dxdy$$

$$=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}dy\cdot\int_{-\infty}^{\infty}(x-m_1)(y-m_2)\cdot\exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x-m_1}{\sigma_1}-r\frac{y-m_2}{\sigma_2}\right)^2\right\}dx.$$

注意到由正态分布的期望可得

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\frac{(x-m_1)}{\sigma_1}e^{-\frac{1}{2(1-r^2)}\cdot\left(\frac{x-m_1}{\sigma_1}-r\frac{y-m_2}{\sigma_2}\right)^2}dx$$

$$=\sigma_1\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}z\exp\left\{-\frac{1}{2(1-r^2)}\left(z-r\frac{y-m_2}{\sigma_2}\right)^2\right\}dz$$

$$=\sigma_1 r\left(\frac{y-m_2}{\sigma_2}\right).\quad\left(\text{其中令 } z=\frac{x-m_1}{\sigma_1}.\right)$$

故再由正态分布的方差可得

$$b_{12}=\frac{\sigma_1 r}{\sigma_2}\int_{-\infty}^{\infty}\frac{(y-m_2)^2}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{(y-m_2)^2}{2\sigma_2^2}}dy=r\sigma_1\sigma_2.$$

故得二维正态分布(ξ,η)的数学期望和协方差矩阵为

$$\left.\begin{aligned}&(E(\xi),E(\eta))=(m_1,m_2),\\&\begin{pmatrix}b_{11}&b_{12}\\b_{21}&b_{22}\end{pmatrix}=\begin{pmatrix}\sigma_1^2&r\sigma_1\sigma_2\\r\sigma_1\sigma_2&\sigma_2^2\end{pmatrix}.\end{aligned}\right\}\tag{3.3.10}$$

因此，类似于一维情形，二维正态分布完全由它的数学期望与协方差矩阵(3.3.10)式所唯一地决定(即它的五个参数 m_1, m_2, σ_1,σ_2,r 完全由 $E(\xi)$, $E(\eta)$, $D(\xi)$, $D(\eta)$, b_{12} 这五个数所唯一确定).

由上面的 $b_{12}=r\sigma_1\sigma_2$ 得

$$r=\frac{b_{12}}{\sigma_1\sigma_2}.\tag{3.3.11}$$

通常称由(3.3.11)式定义的 r 为随机变数 ξ,η 的相关系数.

由例 2.3.1 后面的说明知，对于具有二维正态分布 $N(m_1,\sigma_1;m_2,\sigma_2;r)$ 的随机向量 (ξ,η) 而言，当其中的参数 $r=0$ 时，ξ, η 相互独立. 故(3.3.11)式表明，当 $b_{12}=0$ 时，ξ,η 相互独立.

对于二维正态分布，由于

$$\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)}{\sigma_1}\frac{(y-m_2)}{\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}=\lambda^2$$

(λ 是常数)时，密度函数

$$f(x,y)=\text{常数},$$

故称前者为一等概率椭圆. 现让我们来求 (ξ,η) 落入等概率椭圆的概率. 记此椭圆区域(闭域)为:

$G(\lambda)=$

$$\left\{(x,y):\frac{(x-m_1)^2}{\sigma_1^2}-2r\frac{(x-m_1)}{\sigma_1}\frac{(y-m_2)}{\sigma_2}+\frac{(y-m_2)^2}{\sigma_2^2}\leqslant\lambda^2\right\}.$$

令

$$\begin{cases}\dfrac{x-m_1}{\sigma_1}=\rho\cos\theta,\\[2mm]\dfrac{y-m_2}{\sigma_2}=\rho\sin\theta,\end{cases}$$

$$\left|\frac{\partial(x,y)}{\partial(\rho,\theta)}\right|=\begin{vmatrix}\sigma_1\cos\theta & -\rho\sigma_1\sin\theta\\ \sigma_2\sin\theta & \rho\sigma_2\cos\theta\end{vmatrix}=\rho\sigma_1\sigma_2.$$

从而有：(ξ,η)落入等概率椭圆的概率为

$$P(\lambda)=\iint\limits_{G(\lambda)}f(x,y)\,dxdy$$

$$=\frac{1}{2\pi\sqrt{1-r^2}}\int_0^{2\pi}\int_0^{\left(\frac{\lambda^2}{1-r\sin2\theta}\right)^{\frac{1}{2}}}e^{-\frac{1-r\sin2\theta}{2(1-r^2)}\cdot\rho^2}\rho d\rho d\theta$$

$$=\left(1-e^{-\frac{\lambda^2}{2(1-r^2)}}\right)\cdot\frac{1}{2\pi\sqrt{1-r^2}}\int_0^{2\pi}\frac{1-r^2}{1-r\sin2\theta}d\theta$$

$$=1-e^{-\frac{\lambda^2}{2(1-r^2)}}.$$

其中我们用到了

$$1=P(+\infty)=\frac{1}{2\pi\sqrt{1-r^2}}\int_0^{2\pi}\frac{1-r^2}{1-r\sin2\theta}d\theta.$$

当$r=0$时，(即ξ与η不相关)，我们重新得到了瑞利分布函数$P(\lambda)=1-e^{-\frac{\lambda^2}{2}}$．(对比(2.4.25))．

现在让我们将二维的均匀分布与正态分布推广到 n 维上面去，从而得出两个 n 维随机变数的分布函数的具体例子．

例 3.3.4(n 维均匀分布)　若 n 维随机变数$(\xi_1,\xi_2,\cdots,\xi_n)$的密度函数为

$$f(x_1,\cdots,x_n)=\begin{cases}\prod\limits_{i=1}^{n}\dfrac{1}{(b_i-a_i)}, & 当\ a_i\leqslant x_i\leqslant b_i,\ i=1,\cdots,n;\\ 0, & 其它.\end{cases}\tag{3.3.12}$$

则称$(\xi_1,\cdots,\xi_n)$服从(n 维)均匀分布．类似于例 3.3.2，易证其数学期望(向量)与协方差矩阵分别为：

$$(E\xi_1,\cdots,E\xi_n)=\left(\frac{b_1+a_1}{2},\cdots,\frac{b_n+a_n}{2}\right),$$

$$\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \cdots & \cdots & \cdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} \frac{(b_1-a_1)^2}{12} & & 0 \\ & \ddots & \\ 0 & & \frac{(b_n-a_n)^2}{12} \end{pmatrix}.$$

并且，易知 $\xi_1, \cdots, \xi_n$ 是相互独立的. 因为

$$f(x_1, \cdots, x_n) = f_{\xi_1}(x_1) \cdots f_{\xi_n}(x_n)$$

成立.

为引入 n 维正态分布之密度函数的定义，让我们先来回忆一下正定矩阵之定义：称 $n \times n$ 矩阵 Γ 是正定的，若任取非零 $\lambda \in R^n$（即 λ 为非零 n 维向量），有

$$\lambda' \Gamma \lambda > 0,$$

其中 λ' 是 λ 的转置. 亦即，若 $\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$，则 $\lambda' = (\lambda_1, \cdots, \lambda_n)$.

现引入 n 维正态分布之密度函数如下：

例 3.3.5（n 维正态分布） 设 n 维随机变数 $(\xi_1, \cdots, \xi_n)$ 之密度函数为

$$f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(X-M)' \Sigma^{-1} (X-M)}{2}\right), \tag{3.3.13}$$

$$-\infty < x_i < \infty, \quad i = 1, \cdots, n;$$

其中 $|\Sigma|$ 表矩阵 Σ 之行列式，及

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \cdots & \cdots & \cdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}, \qquad M = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}, \qquad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

$\sigma_{ij}, m_i, i, j = 1, \cdots, n$，都是常数；$\Sigma$ 是对称的正定矩阵，Σ^{-1} 是它的逆矩阵. 则称 $(\xi_1, \cdots, \xi_n)$ 服从 n 维正态分布.

让我们证明：对于 n 维正态密度函数 (3.3.13)，有下列性质

成立：

1° $f(x_1,\cdots,x_n)$是密度函数.

2° $E(\xi_i)=m_i,\quad i=1,2,\cdots,n.$

3° $E((\xi_i-E\xi_i)(\xi_j-E\xi_j))=\sigma_{ij},$

$$i,j=1,2,\cdots,n.$$

故可称M与Σ分别为正态分布之数学期望（向量）与协方差矩阵.

证 1°. 由于Σ为对称的正定矩阵，故据代数知，必存在正交矩阵T，使$T'\Sigma T$成为对角线矩阵，记之为

$$C=T'\Sigma T=\begin{pmatrix} c_{11}^2 & & 0 \\ & \ddots & \\ 0 & & c_{nn}^2 \end{pmatrix}, \text{其中 } c_{ii}^2>0,\ i=1,\cdots,n.$$

从而有

$$\Sigma=TCT',\qquad C^{-1}=T'\Sigma^{-1}T. \tag{3.3.14}$$

作变换

$$X-M=TY.\quad \left(\text{雅可比行列式}\left|\frac{\partial X}{\partial Y}\right|=|T|=1\right)$$

即得（下面记R^n为n维欧氏空间）

$$\int_{R^n}\cdots\int f(x_1,\cdots,x_n)\,dx_1\cdots dx_n$$

$$=\frac{1}{(2\pi)^{n/2}|C|^{1/2}}\int_{R^n}\cdots\int\exp\left(-\frac{Y'C^{-1}Y}{2}\right)dy_1\cdots dy_n$$

$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,c_i}\int_{-\infty}^{\infty}\exp\left(-\frac{y_i^2}{2c_i^2}\right)dy_i=1.$$

证 2° 运用上面的变换，可得

$$E\xi_i=\int_{R^n}\cdots\int x_i\cdot\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{(X-M)'\Sigma^{-1}(X-M)}{2}\right)$$

$$\cdot dx_1\cdots dx_n$$

$$=\int_{R^n}\cdots\int\left(m_i+\sum_{k=1}^{n}t_{ik}y_k\right)\frac{1}{(2\pi)^{n/2}|C|^{1/2}}\exp\left(-\frac{YC^{-1}Y}{2}\right)$$

$\cdot dy_1\cdots dy_n=m_i$.

(其中用到了：t_{ik} 是 T 中第 i 行第 k 列元素，且

$$\int_{-\infty}^{\infty} m_i\cdot\frac{1}{\sqrt{2\pi}\ c_i}\exp\left(-\frac{y_i^2}{2c_i^2}\right)dy_i=m_i,$$

$$\int_{-\infty}^{\infty} y_k\cdot\frac{1}{\sqrt{2\pi}\ c_k}\exp\left(-\frac{y_k^2}{2c_k^2}\right)dy_k=0\Big).$$

证 3°. 同样运用上面的变换，并注意到(3.3.14)，可得

$$b_{ij}=\int\cdots\int_{R^n}(x_i-m_i)(x_j-m_j)f(x_1,\cdots,x_n)dx_1\cdots dx_n$$

$$=\int\cdots\int_{R^n}\left(\sum_{k=1}^{n}t_{ik}y_k\right)\left(\sum_{l=1}^{n}t_{jl}y_l\right)\frac{\exp\left(-\frac{YC^{-1}Y}{2}\right)}{(2\pi)^{n/2}|C|^{1/2}}dy_1\cdots dy_n$$

$$=\int\cdots\int_{R^n}\sum_{k=1}^{n}t_{ik}t_{jk}y_k^2\frac{1}{(2\pi)^{n/2}|C|^{1/2}}\exp\left(-\frac{YC^{-1}Y}{2}\right)dy_1\cdots dy_n$$

$$=\sum_{k=1}^{n}t_{ik}t_{jk}c_k^2=\sigma_{ij}.$$

容易验证：当 $n=2$ 时，(3.3.13) 即是已熟知的二维正态分布.

事实上，在此情形，

$$\Sigma=\begin{pmatrix}b_{11} & b_{12}\\ b_{21} & b_{22}\end{pmatrix},\qquad M=\begin{pmatrix}m_1\\ m_2\end{pmatrix}.$$

故

$$|\Sigma|=b_{11}b_{22}-b_{12}^2=b_{11}b_{22}\left(1-\frac{b_{12}^2}{b_{11}b_{22}}\right)$$

$$=\sigma_1^2\sigma_2^2(1-r^2);$$

$$\Sigma^{-1}=\frac{1}{|\Sigma|}\begin{pmatrix}b_{22} & -b_{21}\\ -b_{12} & b_{11}\end{pmatrix};$$

$$(X-M)'\Sigma^{-1}(X-M)=\frac{1}{|\Sigma|}(b_{22}(x_1-m_1)^2$$

$$-2b_{12}(x-m_1)(x_2-m_2)+b_{11}(x_2-m_2)^2)$$

$$=\left(\frac{(x_1-m_1)^2}{\sigma_1^2}-2r\frac{(x_1-m_1)}{\sigma_1}\frac{(x_2-m_2)}{\sigma_2}\right.$$

$$\left.+\frac{(x_2-m_2)^2}{\sigma_2^2}\right)\Big/(1-r^2)$$

因此，在 $n=2$ 的情形，由(3.3.13)便重新得到例 3.3.3.

当 $n=3$ 时，且设 Σ 为对角线矩阵，由(3.3.13)可得到

$$f(x_1,x_2,x_3)=\frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3}\exp\left(-\frac{1}{2}\sum_{i=1}^{3}\frac{(x_i-m_i)^2}{\sigma_i^2}\right),$$

其中 $\sigma_i=D(\xi_i)$. 上式可解释为分子运动中分子沿坐标轴的分速 ξ_1,ξ_2,ξ_3 的概率分布密度(也称马克斯威分布)，为此，只需取 $\sigma_1^2=\sigma_2^2=\sigma_3^2=\frac{1}{hm}$；其中 m 表示分子的质量，h 是一常数。

§ 3.4 多维随机变数的函数的数字特征

在 § 3.1 中对于一维随机变数的函数的数字特征，定理 3.1.1 给出了它的计算公式. 这一节，我们介绍对于多维随机变数的函数的数字特征. 与一维时一样，有如下定理(这里只作叙述，不予证明).

定理 3.4.1 设 $F(x_1,x_2,\cdots,x_n)$ 为随机向量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的分布函数，$g(x_1,x_2,\cdots,x_n)$ 为 n 维连续函数. 则

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}|g(x_1,x_2,\cdots,x_n)|\,dF(x_1,x_2,\cdots,x_n)<\infty,$$

是 $E[g(\xi_1,\xi_2,\cdots,\xi_n)]$ 存在的充分必要条件. 并且，当它存在时，有

$$E[g(\xi_1,\xi_2,\cdots,\xi_n)]$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,x_2,\cdots,x_n)dF(x_1,x_2,\cdots,x_n). \quad (3.4.1)$$

(3.4.1)及 R-S 积分的性质，立即推知，

$$g(\xi_1,\cdots,\xi_n)\geqslant 0, Eg(\xi_1,\cdots,\xi_n)\text{存在}$$
$$\Rightarrow Eg(\xi_1,\cdots,\xi_n)\geqslant 0.$$

这一性质通常称为：数学期望具有单调性.

现在当 $(\xi_1,\xi_2,\cdots,\xi_n)$ 为离散型随机变数时，有

$$E[g(\xi_1,\xi_2,\cdots,\xi_n)]$$
$$=\sum_{i_1,i_2,\cdots,i_n=1}^{\infty}g(x_{i_1}^{(1)}x_{i_2}^{(2)}\cdots x_{i_n}^{(n)})P(x_{i_1}^{(1)},x_{i_2}^{(2)},\cdots,x_{i_n}^{(n)}),$$

其中 $P(x_{i_1}^{(1)},x_{i_2}^{(2)},\cdots,x_{i_n}^{(n)})=P\{\xi_1=x_{i_1}^{(1)},\xi_2=x_{i_2}^{(2)},\cdots,\xi_n=x_{i_n}^{(n)}\}$.

当 $(\xi_1,\xi_2,\cdots,\xi_n)$ 为连续型随机变数时，有

$$E[g(\xi_1,\xi_2,\cdots,\xi_n)]$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,x_2,\cdots,x_n)f(x_1,x_2,\cdots x_n)dx_1dx_2\cdots dx_n,$$

其中 $f(x_1,x_2,\cdots,x_n)$ 为 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的密度函数.

定理 3.4.1 表明：求 $g(\xi_1,\xi_2,\cdots,\xi_n)$ 的数学期望，不必先求它的分布函数，然后按定义用(3.1.22)式去求．而直接由(3.4.1)式计算即可.

兹举一例说明.

例 3.4.1 试求例 2.4.6(瑞利分布)中之随机变数 η_1 的数学期望.

解 法 1．先求 $\eta_1=\sqrt{\xi_1^2+\xi_2^2}$ 的密度函数．由例 2.4.6 已求得

$$f_{\eta_1}(t)=\begin{cases}\dfrac{t}{\sigma^2}e^{-\frac{t^2}{2\sigma^2}}, & t>0,\\ 0, & \text{当 } t\leqslant 0.\end{cases}$$

故

$$E\eta_1=\int_0^{\infty}\frac{t^2}{\sigma^2}e^{-\frac{t^2}{2\sigma^2}}dt$$

$$=-te^{-\frac{t^2}{2\sigma^2}}\Big|_{t=0}^{t=\infty}+\int_0^{\infty}e^{-\frac{t^2}{2\sigma^2}}dt \quad \text{（分部积分法）}$$

$$=0+\frac{\sqrt{2\pi}}{2}\sigma=\sqrt{\frac{\pi}{2}}\sigma.\text{（由正态分布）。}$$

法 2　运用公式(3.4.1)，可得

$$E\eta_1=\int_{-\infty}^{+\infty}\int_{-\infty}^{\infty}\sqrt{x_1^2+x_2^2}f_{\xi_1,\xi_2}(x_1,x_2)\,dx_1dx_2$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sqrt{x_1^2+x_2^2}(e^{-(x_1^2+x_2^2)/2\sigma^2}/2\pi\sigma^2)\,dx_1dx_2$$

$$=\int_0^{\infty}\int_0^{2\pi}(re^{-\frac{r^2}{2\sigma^2}}/2\pi\sigma^2)\cdot rd\theta dr$$

$$=\int_0^{\infty}\frac{r^2}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}dr=\sqrt{\frac{\pi}{2}}\sigma.$$

比较上面两种计算方法，可知法 2 省去了求密度函数 $f_{\eta_1}(t)$ 之步骤．因此较为简便．

在实际问题中，有时需计算二个随机变数 ξ 与 η 的 $E\max(\xi,\eta)$（例如：最大风速的平均值）；或计算 $E\min(\xi,\eta)$（例如：最小损耗的平均值）．

例 3.4.2　设 ξ 与 η 相互独立，且都服从 $N(0,1)$，求

$$E\min(\xi,\eta).$$

解　由对称性，得

$$E\min(\xi,\eta)=2\int_{-\infty}^{\infty}\int_{-\infty}^{x}y\cdot\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}dy\,dx$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2}}\cdot(-e^{-y^2/2})\Big|_{y=-\infty}^{y=x}dx=-\frac{1}{\pi}\int_{-\infty}^{\infty}e^{-x^2}dx=-\frac{1}{\sqrt{\pi}}.$$

从定理 3.4.1 看到，由(3.3.2)式定义的协方差 b_{ik} 即为

$$b_{ik}=E\{[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\}. \tag{3.4.2}$$

有了定理 3.4.1，我们可以得到有关多维随机变数的函数的

数学期望与方差的性质，在计算多维随机变数的数字特征时屯给我们许多方便之处.

性质 3.4.1 设 n 维随机变数 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的数学期望存在. 则有

1）线性性质：对任意常数 $c_i(i=1,2,\cdots,n)$ 有

$$E\left(\sum_{i=1}^{n} c_i \xi_i\right)=\sum_{i=1}^{n} c_i E(\xi_i). \tag{3.4.3}$$

2）若 $\xi_1, \xi_2, \cdots, \xi_n$ 相互独立，则

$$E\left(\prod_{i=1}^{n} \xi_i\right)=\prod_{i=1}^{n} E(\xi_i). \tag{3.4.4}$$

证 1）由定理3.4.1及 $R\text{-}S$ 积分性质得

$$\begin{aligned}E\left(\sum_{i=1}^{n} c_i \xi_i\right)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\sum_{i=1}^{n} c_i x_i\right)dF(x_1, x_2, \cdots, x_n)\\&=\sum_{i=1}^{n} c_i \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i dF(x_1, x_2, \cdots, x_n)\\&=\sum_{i=1}^{n} c_i E(\xi_i).\end{aligned}$$

2）仅证 $n=2$ 并设 (ξ_1, ξ_2) 为连续型的情形.

设 $f(x_1, x_2)$ 及 $f_1(x_1), f_2(x_2)$ 为 (ξ_1, ξ_2) 及 ξ_1，ξ_2 的密度函数，按定理 3.4.1 及(3.4.3)式，并由 ξ_1, ξ_2 的独立性，有

$$\begin{aligned}E\left(\prod_{i=1}^{2} \xi_i\right)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 dF(x_1, x_2)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2)\, dx_1 dx_2\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f_1(x_1) f_2(x_2)\, dx_1 dx_2\\&=\int_{-\infty}^{\infty} x_1 f_1(x_1)\, dx_1 \cdot \int_{-\infty}^{\infty} x_2 f_2(x_2)\, dx_2\end{aligned}$$

$$=E(\xi_1)\cdot E(\xi_2).$$ ∎

性质 3.4.2 设 c_i 为常数，ξ_i 为随机变数，且 $E(\xi_i^2)<\infty(i=1,2,\cdots,n)$，则

1) $$D\left(\sum_{i=1}^{n}c_i\xi_i\right)=\sum_{i=1}^{n}c_i^2D(\xi_i)+\sum_{\substack{i,k=1\\i\neq k}}^{n}c_ic_kb_{ik},\tag{3.4.5}$$

其中 $b_{ik}=E\{[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\}$.

特别，若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立，则 $b_{ik}=0$（当 $i\neq k$），且

$$D\left(\sum_{i=1}^{n}c_i\xi_i\right)=\sum_{i=1}^{n}c_i^2D(\xi_i).\tag{3.4.6}$$

2) $(E(\xi_1\cdot\xi_2))^2\leqslant E(\xi_1^2)\cdot E(\xi_2^2)$.（许瓦兹不等式）

证 1) 由方差的定义及性质 3.4.1 得

$$D\left(\sum_{i=1}^{n}c_i\xi_i\right)=E\left\{\left[\sum_{i=1}^{n}c_i\xi_i-E\left(\sum_{i=1}^{n}c_i\xi_i\right)\right]^2\right\}$$

$$=E\left\{\sum_{i=1}^{n}c_i[\xi_i-E(\xi_i)]\right\}^2$$

$$=E\left\{\sum_{i=1}^{n}\sum_{k=1}^{n}c_ic_k[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\right\}$$

$$=\sum_{i=1}^{n}\sum_{k=1}^{n}c_ic_kE\{[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\}$$

$$=\sum_{i=1}^{n}c_i^2D(\xi_i)+\sum_{\substack{i,k=1\\i\neq k}}^{n}c_ic_kb_{ik}.$$

特别当 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立时，由性质 3.4.1 之 2)得，当 $i\neq k$ 时

$$b_{ik}=E\{[\xi_i-E(\xi_i)][\xi_k-E(\xi_k)]\}$$

$$=E[\xi_i-E(\xi_i)]\cdot E[\xi_k-E(\xi_k)]$$

$$=[E(\xi_i)-E(\xi_i)][E(\xi_k)-E(\xi_k)]=0.$$

故由(3.4.5)式即得(3.4.6)式.

现证 2）利用 R-S 积分中的许瓦兹不等式即得

$$[E(\xi_1\cdot\xi_2)]^2=\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2dF(x_1,x_2)\right]^2$$
$$\leqslant\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1^2dF(x_1,x_2)\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_2^2dF(x_1,x_2)$$
$$=E(\xi_1^2)\cdot E(\xi_2^2).\ \blacksquare$$

如若用 $\xi_1-E(\xi_1)$ 代替上式中之 ξ_1，用 $\xi_2-E(\xi_2)$ 代替上式中之 ξ_2，则我们又一次得到了性质 3.3.1 之 3）当 $i=1$，$k=2$ 时的情形.

利用上述数学期望与方差的性质，下面计算几个例子.

例 3.4.3 设进行 n 次独立试验，事件 A 在第 k 次试验发生的概率为 p_k，求事件 A 在 n 次试验中出现的总次数 μ 的数学期望与方差.

解 记 μ_k 为第 k 次试验事件 A 发生的次数，则 μ_k 的分布列为

μ_k	0	1
$P(\mu_k=i)$	$1-p_k$	p_k

记 $q_k=1-p_k$，则 $\mu=\mu_1+\mu_2+\cdots+\mu_n$. 由例3.1.2 $E(\mu_k)=p_k$，$D(\mu_k)=p_kq_k$. 故依性质 3.4.1 之1）和性质 3.4.2 得

$$E(\mu)=\sum_{k=1}^{n}p_k,$$
$$D(\mu)=\sum_{k=1}^{n}p_kq_k.$$

特别在当 $p_k=p$ 时的贝努里概型，有

$$E(\mu)=np,\ D(\mu)=npq.$$

或者有

$$E\left(\frac{\mu}{n}\right)=p,\qquad D\left(\frac{\mu}{n}\right)=\frac{pq}{n}.$$

上式$E\left(\frac{\mu}{n}\right)=p$表明：在$n$次独立试验中，事件$A$出现次数为$\mu$，则其平均每次试验出现的次数$\frac{\mu}{n}=\frac{1}{n}\sum_{k=1}^{n}\mu_k$的"期望值"为$p$.

例3.4.4 设对某目标进行射击，每次击发一枚子弹，直到击中n次为止. 设各次射击相互独立，且每次射击时击中目标的概率为p，试求子弹的消耗量ξ的数学期望.

解 记ξ_k为第$k-1$次击中至第k次击中目标之间所消耗的子弹数. 这样，ξ_k可取值1,2,…，其分布列可由下表给出：（记$q=1-p$并设ξ_1为第一次击中目标所消耗的子弹数）

ξ_k	1	2	3	…	n	…
$P\{\xi_k=n\}$	p	qp	q^2p	…	$q^{n-1}p$	…

$k=1,2,\cdots$

（通常称上表为几何分布的分布列）则n次击中目标所需子弹数为

$$\xi=\xi_1+\xi_2+\cdots+\xi_n.$$

由性质3.4.1之1)及$\xi_i(i=1,2,\cdots,n)$同分布得

$$E(\xi)=\sum_{i=1}^{n}E(\xi_i)=nE(\xi_1),$$

其中

$$E(\xi_1)=\sum_{k=1}^{\infty}kq^{k-1}p=p\cdot\left(\frac{1}{1-x}\right)'\bigg|_{x=q}=p\frac{1}{(1-x)^2}\bigg|_{x=q}$$

$$=\frac{p}{(1-q)^2}=\frac{1}{p}.$$

故

$$E(\xi)=\frac{n}{p}.$$

上式说明，耗弹多少与击中概率p成反比，这是很符合直观的.

例3.4.5 设在同一组条件下独立地对某物的长度a进行了

n 次测量．第 k 次测量的结果为 ξ_k，它是随机变数．又设所有的 ξ_k 服从 $N(a,\sigma)$，试计算 n 次测量结果的平均长度 $\frac{1}{n}\sum_{k=1}^{n}\xi_k$ 的数学期望和方差．

解 由性质 3.4.1 及性质 3.4.2 得

$$E\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n}\sum_{k=1}^{n}E(\xi_k)=a. \tag{3.4.7}$$

又

$$D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D(\xi_k)=\frac{1}{n}\sigma^2. \tag{3.4.8}$$

(3.4.7)式表明，n 次测量结果的平均值的"期望"恰好是物体长度 a，而(3.4.8)式则表明，n 次测量结果的平均值所产生的离散程度（或说绝对误差）比一次测量的离散程度（或说误差）来得小．因此，在实际的测量工作中常常利用这一结果，以减少误差．

我们在 § 3.3 介绍具有二维正态 $N(m_1,\sigma_1;m_2,\sigma_2;r)$ 的随机变数 (ξ_1,ξ_2) 时，得到了(3.3.11)式，其中参数 $r=\frac{b_{12}}{\sigma_1\sigma_2}$，而且在第二章中得到过当 $r=0$ 时，则 ξ_1，ξ_2 相互独立．另一方面，我们从(3.4.2)式知，$b_{12}=E\{[\xi_1-E(\xi_1)][\xi_2-E(\xi_2)]\}$．因此，由性质 3.4.2 知，当 ξ_1、ξ_2 独立时，$b_{12}=0$．从而 $r=0$．综上所述，我们得到了二维正态分布的一个重要性质．

性质 3.4.3 若随机变数 $(\xi_1,\ \xi_2)$ 服从二维正态 $N(m_1,\ \sigma_1;m_2,\sigma_2;r)$，则 ξ_1 与 ξ_2 相互独立的充分必要条件是 $r=0$．

对于具有二维正态分布的随机变数 $(\xi_1,\ \xi_2)$，其参数 r 的表达式

$$r=\frac{E\{[\xi_1-E(\xi_1)][\xi_2-E(\xi_2)]\}}{\sqrt{D(\xi_1)}\sqrt{D(\xi_2)}}$$

当 $r=0$ 时，刻划了随机变数 ξ_1 与 ξ_2 的独立性这一特性. 但对一般的随机变数，是否仍可用上述表达式来作为研究随机变数之间某种性质的数字特征呢？为此，我们引入如下定义：

定义 3.4.1 设随机变数 ξ_1,ξ_2 的方差 $D(\xi_1)$，$D(\xi_2)$ 存在且均大于 0. 记

$$\rho_{12}=\frac{E\{[\xi_1-E(\xi_1)][\xi_2-E(\xi_2)]\}}{\sqrt{D(\xi_1)}\sqrt{D(\xi_2)}}, \tag{3.4.9}$$

称 ρ_{12} 为 ξ_1 与 ξ_2 的相关系数. 在不致产生混乱时，简记为 ρ.

利用定义 3.3.1 的写法，(3.4.9) 式可写为

$$\rho_{12}=\frac{b_{12}}{\sqrt{b_{11}}\sqrt{b_{22}}}. \tag{3.4.10}$$

一般的有

$$\rho_{ik}=\frac{b_{ik}}{\sqrt{b_{ii}}\sqrt{b_{kk}}}. \tag{3.4.11}$$

并称矩阵

$$R=\begin{pmatrix}\rho_{11} & \rho_{12} & \cdots & \rho_{1n}\\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn}\end{pmatrix} \tag{3.4.12}$$

为相关矩阵. 若记

$$C=\begin{pmatrix}\frac{1}{\sqrt{D(\xi_1)}} & & 0\\ & \ddots & \\ 0 & & \frac{1}{\sqrt{D(\xi_n)}}\end{pmatrix}\text{（对角线矩阵）} \tag{3.4.13}$$

则由 (3.3.3) 及上面，得 $C\Sigma C=R$.

应用定义 3.4.1，则性质 3.4.3 可叙述为：对服从二维正态 $N(m_1,\sigma_1;m_2,\sigma_2;r)$ 的随机变数 (ξ_1,ξ_2)，ξ_1 与 ξ_2 独立的充分必要

条件为它们的相关系数为 0 (或说不相关).

性质 3.4.4 随机变数 ξ_1 与 ξ_2 的相关系数 ρ 满足下列性质:

1° $-1\leqslant\rho\leqslant1$;

2° $|\rho|=1$ 的充分必要条件为 ξ_1 与 ξ_2 以概率为 1 线性相关. 即存在常数 α,β, 且 $\alpha\neq0$, 使得

$$P\{\xi_2=\alpha\xi_1+\beta\}=1.$$

证 1° 由性质 3.3.1, 即得

$$|\rho|^2=\frac{|E\{[\xi_1-E(\xi_1)][\xi_2-E(\xi_2)]\}|^2}{D(\xi_1)\cdot D(\xi_2)}$$

$$\leqslant\frac{E[\xi_1-E(\xi_1)]^2\cdot E[\xi_2-E(\xi_2)]^2}{D(\xi_1)\cdot D(\xi_2)}=1.$$

现证 2° 先证必要性:

$$\because\quad D\left(\frac{\xi_1}{\sqrt{b_{11}}}\pm\frac{\xi_2}{\sqrt{b_{22}}}\right)=\frac{b_{11}}{b_{11}}+\frac{b_{22}}{b_{22}}\pm2\frac{b_{12}}{\sqrt{b_{11}}\sqrt{b_{22}}}=2(1\pm\rho),$$

当 $\rho=1$ 时, 用上式: $D\left(\frac{\xi_1}{\sqrt{b_{11}}}-\frac{\xi_2}{\sqrt{b_{22}}}\right)=2(1-\rho)$ 得

$$D\left(\frac{\xi_1}{\sqrt{b_{11}}}-\frac{\xi_2}{\sqrt{b_{22}}}\right)=0.$$

由定理 3.1.2 的推论知

$$P\left\{\frac{\xi_1}{\sqrt{b_{11}}}-\frac{\xi_2}{\sqrt{b_{22}}}=\frac{E(\xi_1)}{\sqrt{b_{11}}}-\frac{E(\xi_2)}{\sqrt{b_{22}}}\right\}=1,$$

即有

$$P\{\xi_2=\alpha\xi_1+\beta\}=1.$$

对于 $\rho=-1$, 可仿上述方法证明.

下面证明充分性: 设有常数 α, β, 且 $\alpha>0$, 使得 $\xi_2=\alpha\xi_1+\beta$. 则有

$$\begin{aligned}b_{12}&=E[\xi_1-E(\xi_1)][\xi_2-E(\xi_2)]\\&=E[\xi_1-E(\xi_1)][\alpha\xi_1+\beta-\alpha E(\xi_1)-\beta]\\&=\alpha D(\xi_1).\end{aligned}$$

同理

$$b_{22}=E[\xi_2-E(\xi_2)]^2=\alpha^2D(\xi_1).$$

由此得

$$\rho=\frac{b_{12}}{\sqrt{b_{11}b_{22}}}=\frac{\alpha}{|\alpha|}.$$

若 $\alpha>0$，则 $\rho=1$；若 $\alpha<0$，则 $\rho=-1$. ∎

由此可以看到，数字特征 ρ 描述了随机变数 ξ_1,ξ_2 的线性相关程度，因而当 $\rho=0$ 时，我们自然称 ξ_1，ξ_2 不相关.（注意，我们这里指的是它们之间没有线性相关关系.）

不相关与相互独立性，在一般情形是不等价的. 诚然，当 ξ_1 与 ξ_2 相互独立时，可推得 ξ_1 与 ξ_2 不相关（即线性不相关），但反过来则不一定正确.

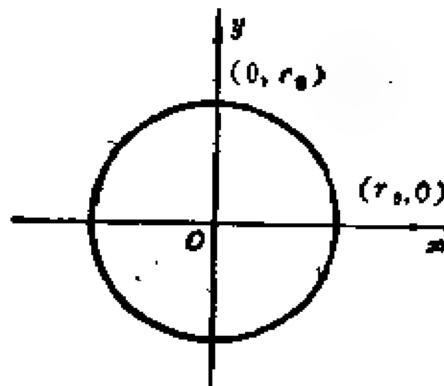


图 3.4.1

例3.4.6 设 (ξ,η) 均匀分布于以坐标原点为中心，r_0 为半径的圆的内部（见图 3.4.1），则 $E(\xi)=0$，$E(\eta)=0$，$E(\xi\eta)=0$，但 ξ，η 不独立.

证 因为 (ξ,η) 为均匀分布，故 (ξ,η) 之密度函数为

$$f(x,y)=\begin{cases}a, & \text{当 } x^2+y^2\leqslant r_0^2;\\ 0, & \text{当 } x^2+y^2>r_0^2.\end{cases}$$

由于

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dx\,dy=\iint\limits_{x^2+y^2\leqslant r_0^2}a\,dx\,dy=a\iint\limits_{x^2+y^2\leqslant r_0^2}dx\,dy,$$

而圆面积等于 πr_0^2，故得 $a=\frac{1}{\pi r_0^2}$. 注意到

$$f_\xi(x)=\int_{-\sqrt{r_0^2-x^2}}^{\sqrt{r_0^2-x^2}}\frac{1}{\pi r_0^2}dy=\frac{2\sqrt{r_0^2-x^2}}{\pi r_0^2},$$

$$f_\eta(y)=\int_{-\sqrt{r_0^2-y^2}}^{\sqrt{r_0^2-y^2}}\frac{1}{\pi r_0^2}dx=\frac{2\sqrt{r_0^2-y^2}}{\pi r_0^2},$$

显然

$$f(x,y)\neq f_\xi(x)\cdot f_\eta(y),$$

故 ξ,η 不是独立的. 但

$$E(\xi)=\int_{-\infty}^{\infty}xf_\xi(x)\,dx=\int_{-r_0}^{r_0}\frac{2x\sqrt{r_0^2-x^2}}{\pi r_0^2}dx=0.$$

($\because$被积函数是奇函数)同理可得, $E(\eta)=0$.

又由于 $f(x,y)$ 的对称性, 可知

$$E\{[\xi-E(\xi)]\cdot[\eta-E(\eta)]\}=\iint\limits_{x^2+y^2\leq r_0^2}\frac{x\cdot y}{\pi r_0^2}dx\,dy=0.$$

从而 ξ,η 的相关系数为 0, 但 ξ,η 不是独立的. ∎

性质 3.4.5 设随机变数 ξ_1,ξ_2 的方差存在. 令

$$\xi=a_1\xi_1+b_1,$$

$$\eta=a_2\xi_2+b_2,$$

则 ξ,η 的相关系数 $\rho_{\xi\eta}=\frac{a_1a_2}{|a_1a_2|}\rho_{12}$, 其中 ρ_{12} 为 ξ_1, ξ_2 的相关系数.

证 $\because$ ξ 与 η 的协方差和方差为

$$\begin{aligned}b_{\xi\eta}=&E(\{a_1\xi_1+b_1-[a_1E(\xi_1)+b_1]\}\\&\{a_2\xi_2+b_2-[a_2E(\xi_2)+b_2]\})\\=&a_1a_2b_{12},\end{aligned}$$

$$D(\xi)=a_1^2D(\xi_1)=a_1^2b_{11},$$

$$D(\eta)=a_2^2D(\xi_2)=a_2^2b_{22}.$$

由此得

$$\rho_{\xi\eta}=\frac{a_1a_2b_{12}}{\sqrt{a_1^2b_{11}}\sqrt{a_2^2b_{22}}}=\frac{a_1a_2}{|a_1a_2|}\rho_{12}.$$ ∎

由性质 3.4.5 可以得到，特别当

$$\xi=\frac{\xi_1-E(\xi_1)}{\sqrt{D(\xi_1)}},\qquad \eta=\frac{\xi_2-E(\xi_2)}{\sqrt{D(\xi_2)}}$$

时，由于

$$E(\xi)=E(\eta)=0, D(\xi)=D(\eta)=1, b_{\xi\eta}=\frac{b_{12}}{\sqrt{D(\xi_1)D(\xi_2)}}=\rho_{12},$$

得到
$$b_{\xi\eta}=\rho_{\xi\eta}=\rho_{12}.$$

性质 3.4.5 表明，相关系数不依赖于原点和单位的选取，正是这个缘故，所以我们用相关系数而不用协方差来作为刻划随机变数间相互关系（依线性关系）的密切程度的一个数字特征. 诚然，协方差也可作为上述的一个数字特征. 这一点，我们将在随机过程那一章里叙述.

现在，让我们来指出一个重要的事实：类似于二维情形那样，关于 n 维正态分布，我们也可以有独立性与线性无关等价的结论. 写成下面的定理.

定理 3.4.2 设 $(\xi_1,\cdots,\xi_n)$ 服从 n 维正态分布，且 $D(\xi_i)>0$，$i=1,2,\cdots,n$. 则下列命题等价：

1° $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立；

2° $\xi_1,\xi_2,\cdots,\xi_n$ 线性无关；即

$\rho_{ij}=0$，当 $i\neq j$; $i,j=1,2,\cdots,n$;

3° 协方差矩阵是对角矩阵；即

$b_{ij}=0$，当 $i\neq j$, $i,j=1,\cdots,n$.

证 1°⇒3° 是显然的. 因为，由独立性可得

$$b_{ij}=E((\xi_i-E\xi_i)\cdot(\xi_j-E\xi_j))=(E\xi_i-E\xi_i)\cdot(E\xi_j-E\xi_j)$$
$$=0.\ (\text{当 } i\neq j)$$

2°⟺3°. 由下面等式即知此等价性：

$$\rho_{ij}=b_{ij}/(\sqrt{D(\xi_i)D(\xi_j)}).$$

$3°\Rightarrow 1°$. 由 $3°$,

$$\Sigma=\begin{pmatrix} b_{11} & & 0 \\ & \ddots & \\ 0 & & b_{nn} \end{pmatrix}.$$

故由(3.3.13)知$(\xi_1,\cdots,\xi_n)$之密度函数为

$$f(x_1,\cdots,x_n)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi b_{ii}}}\exp\left(-\frac{(x_i-m_i)^2}{2b_{ii}}\right)$$

$$=\prod_{i=1}^{n}f_{\xi_i}(x_i).$$

亦即 $\xi_1,\cdots,\xi_n$ 为相互独立. ▌

推论 3.4.1 若 $(\xi_1,\cdots,\xi_{n-1})$ 服从 $n-1$ 维正态分布，且 $D(\xi_i)>0$, $i=1,2,\cdots,n-1$. 又设 ξ_n 是一个随机变数，满足 $D(\xi_n)=0$.(此时，称$(\xi_1,\cdots,\xi_n)$ 服从一个退化为 $n-1$ 维的正态分布). 则下列命题等价:

$1°$ $\xi_1,\cdots,\xi_{n-1},\xi_n$ 相互独立;

$2°$ $(\xi_1,\cdots,\xi_{n-1})$之协方差矩阵为对角线矩阵;

$3°$ $\xi_1,\cdots,\xi_{n-1}$ 线性无关.

证 仅需证 $2°\Rightarrow 1°$. 首先运用定理 3.4.2, 已知 $\xi_1,\cdots,\xi_{n-1}$ 相互独立. 现由车贝谢夫不等式(定理 3.1.2) 之推论，知概率为 1 地有

$$\xi_n=E(\xi_n). \ (\text{常数}).$$

但概率为 1 的常数随机变数(可简称为常数随机变数)必满足

$$P(\xi_n<x_n)=\begin{cases}0, \text{当 } x_n\leqslant E\xi_n,\\ 1, \text{当 } x_n>E\xi_n.\end{cases}$$

故当 $x_n\leqslant E\xi_n$ 时，有

$$P(\xi_1<x_1,\cdots,\xi_{n-1}<x_{n-1},\xi_n<x_n)=0$$

$$=P(\xi<x_1)\cdots P(\xi_{n-1}<x_{n-1})\cdot P(\xi_n<x_n).$$

当 $x_n > E\xi_n$ 时，有

$$
\begin{aligned}
& P(\xi_1 < x_1, \cdots, \xi_{n-1} < x_{n-1}, \xi_n < x_n) \\
= & P(\xi_1 < x_1, \cdots, \xi_{n-1} < x_{n-1}) \cdot P(\xi_n < x_n) \\
= & P(\xi_1 < x_1) \cdots P(\xi_{n-1} < x_{n-1}) \cdot P(\xi_n < x_n).
\end{aligned}
$$

因此，$\xi_1, \cdots, \xi_{n-1}, \xi_n$ 相互独立. ▮

在本节的最后，让我们将已经得到的一些有用的概率不等式归纳一下，以便于查阅. 它们在前面有的已经应用过，有的在后面还将要应用.

车贝谢夫不等式(即定理 3.1.2)　若 ξ 的方差 $D(\xi)$ 存在，则对任意 $e>0$，有

$$P(|\xi - E\xi| \geqslant e) \leqslant D(\xi)/e^2.$$

马尔科夫不等式(即定理 3.2.1)　若 ξ 的 k 阶原点绝对矩 α_k 存在，则对任给 $e>0$，有

$$P(|\xi| \geqslant e) \leqslant \alpha_k / e^k.$$

将车贝谢夫不等式推广，可以得到柯尔莫哥洛夫不等式，它在第五章极限定理中将得到证明.

柯尔莫哥洛夫不等式. 设 $\xi_1, \cdots, \xi_n$ 为相互独立随机变数，且 $E\xi_i = 0, D(\xi_i) < \infty, i = 1, \cdots, n$. 记

$$\xi = \max\left\{\sum_{i=1}^{k} \xi_i; k = 1, 2, \cdots, n\right\}.$$

则对任给 $e>0$，必有

$$P(|\xi| \geqslant e) \leqslant \sum_{i=1}^{n} D(\xi_i)/e^2.$$

许瓦兹不等式(即性质 3.4.2 之 2)　设 $\xi_i, i=1,2$，皆有有限的二阶矩，亦即 $E\xi_i^2 < \infty, i=1,2$，则有

$$[E(\xi_1 \cdot \xi_2)]^2 \leqslant E(\xi_1^2) \cdot E(\xi_2^2).$$

§3.5 条件数学期望

在§2.3中，我们讨论了条件分布函数，给出了在条件 $\xi=x$ 下,随机变数 η 的条件分布函数 $F(y/x)=P\{\eta<y|\xi=x\}$ 以及在条件 $\eta=y$ 下，随机变数 ξ 的条件分布函数 $F(x/y)=P\{\xi<x|\eta=y\}$ 的定义．仿照§3.1,对于上述条件分布函数，我们引入条件数学期望的定义,平行地列出有关性质,并且对条件数学期望的某些特有的性质作简要叙述.

定义 3.5.1 对条件分布函数 $F(y/x)$ 及 $F(x/y)$,若

$$\int_{-\infty}^{\infty}|y|dF(y/x)<\infty,\quad \int_{-\infty}^{\infty}|x|dF(x/y)<\infty,$$

记

$$E(\eta/x)=\int_{-\infty}^{\infty}ydF(y/x),\tag{3.5.1}$$

称 $E(\eta/x)$ 为 $\xi=x$ 条件下，η 的条件数学期望．又记

$$E(\xi/y)=\int_{-\infty}^{\infty}xdF(x/y),\tag{3.5.2}$$

称 $E(\xi/y)$ 为 $\eta=y$ 条件下,ξ 的条件数学期望.

从(3.5.1)及(3.5.2)两式看到，$E(\eta/x)$ 依赖于值 x,若记为 $g(x)$,它依随机变数 ξ 的取值不同而取不同的值．可证 $g(\xi)=E(\eta/\xi)$ 是一个随机变数．同理,$E(\xi/\eta)$ 也是一个随机变数.

类似于(3.1.23)式,若 (ξ,η) 为离散型随机变数，其概率函数 $p(i,k)=P\{\xi=x_i,\eta=y_k\}$, $i,k=1,2,\cdots$．条件概率函数 $p(k/i)=P\{\eta=y_k|\xi=x_i\}$ 及 $p(i/k)=P\{\xi=x_i|\eta=y_k\}$, $i,k=1,2,\cdots$．则有

$$\left.\begin{aligned}
E(\eta/x_i)&=\sum_k y_k p(k/i)=\sum_k y_k\frac{p(i,k)}{p(i,\cdot)}\\
&\qquad i=1,2,\cdots,\\
E(\xi/y_k)&=\sum_i x_i p(i/k)=\sum_i x_i\frac{p(i,k)}{p(\cdot,k)}\\
&\qquad k=1,2,\cdots
\end{aligned}\right\}\tag{3.5.3}$$

类似(3.1.24)式，若(ξ,η)为连续型随机变数，其密度函数为$f(x,y)$，条件密度为$f(x/y)$和$f(y/x)$，则有

$$\left.\begin{aligned}E(\eta/x)&=\int_{-\infty}^{\infty}yf(y/x)dy=\int_{-\infty}^{\infty}y\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)dy}dy\\&=\frac{\int_{-\infty}^{\infty}yf(x,y)dy}{\int_{-\infty}^{\infty}f(x,y)dy},\\E(\xi/y)&=\int_{-\infty}^{\infty}xf(x/y)dx=\int_{-\infty}^{\infty}x\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)dx}dx\\&=\frac{\int_{-\infty}^{\infty}xf(x,y)dx}{\int_{-\infty}^{\infty}f(x,y)dx}.\end{aligned}\right\}\tag{3.5.4}$$

例3.5.1 考虑例2.2.5，由§2.3的"条件分布"一段已计算得

$$p(k/i)=p(i/k)=\frac{1}{n}$$

由此，按(3.5.3)式计算得

$$\begin{aligned}E(\eta/i)&=\sum_{k=1}^{n}kp(k/i)=\frac{1}{n}\sum_{k=1}^{n}k=\frac{1}{n}\left[\frac{n(n+1)}{2}\right]\\&=\frac{n+1}{2}\qquad(i=1,2,\cdots,n),\\E(\xi/k)&=\sum_{i=1}^{n}ip(i/k)=\frac{1}{n}\sum_{i=1}^{n}i=\frac{1}{n}\left[\frac{n(n+1)}{2}\right]\\&=\frac{n+1}{2}\qquad(k=1,2,\cdots,n).\end{aligned}$$

例3.5.2 考虑例2.3.2，我们已求得

$$\left.\begin{aligned}f(y/x)&=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}e^{-\frac{1}{2}\left(\frac{y-rx}{\sqrt{1-r^2}}\right)^2},\\f(x/y)&=\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}e^{-\frac{1}{2}\left(\frac{x-ry}{\sqrt{1-r^2}}\right)^2}.\end{aligned}\right\}\tag{3.5.5}$$

由(3.5.4)计算得

$$E(\eta/x)=\int_{-\infty}^{\infty}yf(y/x)dy$$
$$=\int_{-\infty}^{\infty}y\frac{1}{\sqrt{2\pi}\sqrt{1-r^2}}e^{-\frac{1}{2}\left(\frac{y-rx}{\sqrt{1-r^2}}\right)^2}dy=rx.$$

同理可得

$$E(\xi/y)=ry.$$

事实上，上述计算的结果，我们可由(3.5.5)直接看出. 因为(3.5.5)式的$f(y/x)$及$f(x/y)$为服从正态$N(rx,\sqrt{1-r^2})$和$N(ry,\sqrt{1-r^2})$的密度函数，故有$E(\eta/x)=rx, E(\xi/y)=ry$.

类似定理3.1.1，我们有如下定理.

定理 3.5.1 设$g(x)$于R_1连续，若

$$\int_{-\infty}^{\infty}|g(x)|dF(x/y)<\infty,$$
$$\int_{-\infty}^{\infty}|g(y)|dF(y/x)<\infty.$$

则有

$$\left.\begin{aligned}E(g(\xi)/y)&=\int_{-\infty}^{\infty}g(x)dF(x/y),\\ E(g(\eta)/x)&=\int_{-\infty}^{\infty}g(y)dF(y/x).\end{aligned}\right\}\qquad(3.5.6)$$

***定理 3.5.2** 设ξ,η,ζ皆为随机变数，$g(x)$于R_1连续，且$E(\xi),E(\eta),E(\zeta)$及$E[g(\eta)\cdot\xi]$皆存在. 则

1° 当ξ与η相互独立时，$E(\xi/\eta)=E(\xi)$；

2° $E(\xi)=E[E(\xi/\eta)]$；

3° $E[g(\eta)\cdot\xi/\eta]=g(\eta)\cdot E(\xi/\eta)$；

4° $E[g(\eta)\cdot\xi]=E[g(\eta)\cdot E(\xi/\eta)]$；

5° $E(c/\eta)=c$，c为常数；

6° $E[g(\eta)/\eta]=g(\eta)$；

7° $E[(a\xi+b\eta/\zeta]=aE(\xi/\zeta)+bE(\eta/\zeta)$； ($a,b$为常数)

$8°$ $E[\xi-E(\xi/\eta)]^2 \leqslant E[\xi-g(\eta)]^2$.

证 证 $1°$：由 § 2.3 已得，当 ξ 与 η 相互独立时，有

$$F(x/y)=F_\xi(x).$$

由(3.5.2)立得，对任意 y

$$E(\xi/y)=\int_{-\infty}^{\infty} x dF(x/y)=\int_{-\infty}^{\infty} x dF_\xi(x)=E(\xi).$$

证 $2°$：只证(ξ,η)为连续型随机变数时的情形.

$$\begin{aligned}
E[E(\xi/\eta)]&=\int_{-\infty}^{\infty} E(\xi/y) f_\eta(y)dy\\
&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} xf(x/y)dx\right)f_\eta(y)dy\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x/y)f_\eta(y)dxdy\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy\\
&=\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f(x,y)dy\right)dx\\
&=\int_{-\infty}^{\infty} xf_\xi(x)dx=E(\xi).
\end{aligned}$$

证 $3°$：只须验证，对任意固定的 y，有

$$E[(gy)\cdot\xi/y]=g(y)E(\xi/y)$$

成立即可.

事实上，由(3.5.2)式，有

$$\begin{aligned}
E[g(y)\cdot\xi/y]&=\int_{-\infty}^{\infty} g(y)\cdot xdF(x/y)\\
&=g(y)\int_{-\infty}^{\infty} xdF(x/y)=g(y)E(\xi/y).
\end{aligned}$$

证 $4°$：事实上，由 $2°$ 及 $3°$ 即得.

证 $5°$：由(3.5.2)式即得.

证 $6°$：可由 $3°$ 及 $5°$ 得.

证 $7°$：利用条件联合分布函数及条件期望定义可得.

证 8°：即要证，对任一固定的 y，当 $g(y)=E(\xi/\eta=y)$ 时，$E[\xi-g(\eta)]^2$ 为最小．今就连续型的情形，证明叙述如下：

$$E[\xi-g(\eta)]^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x-g(y)]^2f(x,y)\,dxdy$$

$$=\int_{-\infty}^{\infty}f_\eta(y)\left(\int_{-\infty}^{\infty}[x-g(y)]^2f(x/y)\,dx\right)dy$$

由性质 3.1.4 知道，当 $g(y)=E(\xi/\eta=y)$时，积分

$$\int_{-\infty}^{\infty}[x-g(y)]^2f(x/y)\,dx$$

达到最小，因而当 $g(\eta)=E(\xi/\eta)$时 $E[\xi-g(\eta)]^2$ 为最小．▌

我们进一步考虑 2°，当 η 只取有限个值 $y_i\,(i=1,2,\cdots,n)$，且

$$\sum_{i=1}^{n}P\{\eta=y_i\}=1,\text{其中 } y_i\neq y_j\ \ (i\neq j).$$

若记

$$A_i=\{\eta=y_i\},E(\xi/A_i)=E(\xi/y_i),$$

则有

$$E(\xi)=\sum_{i=1}^{n}E(\xi/A_i)P(A_i). \tag{3.5.7}$$

我们把(3.5.7)式称为全数学期望公式(类似于全概率公式)．

习　题

1. 一个有 n 个钥匙的人要开他的门，他随机而独立地用钥匙试开．分别如下两种情形求试开次数的数学期望和方差：a) 如果试开不成功的钥匙没有从以后的选取中除去；b) 如果除去试开不成功的钥匙．

2. 如果随机变数 ξ 有几何分布：

$$P(\xi=k)=pq^k,\quad(k=0,1,\cdots).$$

证明：$D(\xi)=qp^{-2}$．

3. 设 $P(\xi=n)=\frac{1}{2^n}$，$n=1,2,\cdots$，求 $E(\xi)$ 及 $D(\xi)$．

4. 设随机变数 ξ 取任意正整数的概率像几何数列减少．试选择数列的

首项 a 及公比 q，使得随机变数 ξ 的期望等于 10．并且计算在此条件下，$\{\xi\leqslant 10\}$的概率．

5．无线电台发出的呼叫信号被另一电台接收的概率为 0.2，信号每隔 5 秒钟拍发一次，直到收到对方的回答信号时为止．发出信号和收到信号之间要经过 16 秒钟的时间．求在双方建立起联系以前已拍发的呼唤信号的平均次数．

6．设随机变数 ξ 服从指数分布：

$$f(x)=\begin{cases}be^{-bx}, & 当\ x\geqslant 0;\\ 0, & 当\ x<0,\end{cases}\quad (b>0),$$

求 $E(\xi)$ 及 $D(\xi)$．

7．使仪器停止工作的元件故障数 ξ 的概率分布函数为$F(x)=1-e^{-ax}$ $(x\geqslant 0, a>0)$，当$x<0$时，$F(x)=0$，求 $E(\xi)$ 及 $D(\xi)$．

8．设随机变数 ξ 的分布函数为

$$(x)=\begin{cases}0, & 当\ x<-1;\\ a+b\arcsin x, & 当\ -1\leqslant x<1;\\ 1, & 当\ 1\leqslant x,\end{cases}$$

求 a, b 及 $E(\xi)$、$D(\xi)$．

(β-分布)随机变数 ξ 的密度函数为

$$f(x)=\begin{cases}Ax^{\alpha-1}(1-x)^{\beta-1}, & 当\ 0<x<1;\\ 0, & 其它,\end{cases}$$

其中 $\alpha>0, \beta>0$ 都是常数．求：i) 系数 A；ii) $E(\xi)$及 $D(\xi)$．

10．设 ξ 服从瑞利分布：

$$f(x)=\begin{cases}0, & 当\ x<0;\\ \dfrac{x}{a}e^{-\frac{x^2}{2a}}, & 当\ x\geqslant 0.\end{cases}\quad (a>0),$$

求 $E(\xi)$ 及 $D(\xi)$．

11．随机变数 ξ 的密度函数为

$$f(x)=\frac{1}{2}e^{-|x|},\quad -\infty<x<\infty,$$

求 $E(\xi)$ 及 $D(\xi)$．

12．气体分子的速度服从麦克斯威尔分布

$$f(x)=\begin{cases}Ax^2e^{-\frac{x^2}{a^2}}, & \text{当 } x\geqslant 0;\\ 0, & \text{当 } x<0,\end{cases}$$

其中 $a>0$ 为常数．求：i）系数 A；ii）气体分子速度的均值及方差．

13．试问：联接以 R 为半径的圆周上一已知点与圆周上任意点的弦长的数学期望是多少？

14．如果放射性物质在初始瞬间的质量为 m_0，而在单位时间内任一原子进行核分裂的概率为一常数 p．试求经过 t 时刻后，其质量的数学期望．

15．不动点 O 以高度 h 位于一长为 l 的水平线段 AK 的端点 A 的上方．在 AK 线段上有一随机点 x，它在 AK 线段的所有的位置上等概率，求 OA 与 Ox 线间夹角 φ 的期望．

16．设 μ 是事件 A 在 n 次重复独立试验中的出现次数，在每次试验中 $P(A)=p$．再设随机变数 η 视 μ 为偶数或奇数而取 0 或 1，求 $E(\eta)$．

17．试验成功的概率为 p，不成功的概率为 $1-p$，经过 m 次成功试验后达到预定结果（某事件发生）的概率（条件概率）$G(m)$ 等于

$$G(m)=1-\left(1-\frac{1}{\omega}\right)^m \quad (\omega \text{ 为正常数})．$$

求为了达到预定结果，必须进行的独立试验次数的期望值．

18．设 r，v 均为正常数，若随机变数 ξ 的分布律为

$$P(\xi=k)=(-1)^kC_{-r}^{k}p^k(1-p)^r,\ k=0,1,2,\cdots$$

其中 $0<p<1$，则称 ξ 服从负二项分布．求 $E(\xi)$ 及 $D(\xi)$．

19．设 $\phi(x)>0$，且当 $t>0$ 时是单调上升函数．又设 $E[\phi(|\xi|)]=M$ 存在．证明：对任意 $t>0$，有

$$P\{|\xi|\geqslant t\}\leqslant M/\phi(t)．$$

20．求证：若 $E(e^{a\xi})<\infty$，$(a>0)$ 则

$$P(\xi\geqslant x)\leqslant e^{-ax}\cdot E(e^{a\xi})．$$

21．证明事件在一次试验中发生的次数的方差不超过 1/4．

22．证明：对取值于区间 (a,b) 内的随机变数 ξ，恒成立不等式：

$$a\leqslant E(\xi)\leqslant b;\quad D(\xi)\leqslant\frac{(b-a)^2}{4}．$$

23．设随机变数 ξ 服从 $N(a,\sigma)$，试求 $E|\xi-a|$．

24．如果随机变数 ξ 有分布函数 $F(x)$，证明：当 c 为 ξ 的中位数时，

$E|\xi-c|$（c 为任意常数）达到最小值$\left(\xi\text{的中位数即方程 } F(x)=\frac{1}{2}\text{ 的根}\right)$.

25. 设 k 为正整数且 $E|\xi|^k<\infty$，证明 $E|\xi|^r<\infty$.（$1\leqslant r<k$ 为正整数）.

26. 设 ξ 服从二项分布 $b(n;\ p)$. 证明

$$E(\xi-np)^4=npq(p^3+q^3)+3p^2q^2(n^2-n).$$

27. 在贝努里试验序列中，令 ξ 表示由第一个试验开始算起的游程①的长度(或正面或反面游程). 求 ξ 的分布及 $E(\xi)$、$D(\xi)$.

28. 继上题. 设 η 是第二个游程的长度，求 η 的分布、$E(\eta)$ 及 $D(\eta)$，并求 ξ 和 η 的联合分布.

29. 设随机变数(ξ,η)的联合分布为

$$P(\xi=k_1,\eta=k_2)=\frac{r!}{k_1!k_2!(r-k_1-k_2)!}p_1^{k_1}p_2^{k_2}p_3^{r-k_1-k_2}$$

所定义的三项分布，其中 $p_1+p_2+p_3=1$. 按下述方法求 $E(\xi)$、$D(\xi)$ 及 ξ 和 η 的协方差：

a) 由直接计算；

b) 把 ξ 和 η 都表为 r 个相互独立的随机变数的和.

30. 设随机变数 (ξ,η) 服从二维正态分布：$E(\xi)=E(\eta)=0$，$D(\xi)=D(\eta)=1$，$\rho=R$. 试证：

$$E[\max(\xi,\eta)]=\sqrt{\frac{1-R}{\pi}}.$$

31. 在长为 l 的线段上任意取两点. 求：i) 两点间距离的期望及方差；ii) 求两点间距离的 n 次方的期望及方差.

32. 在 10 件产品中有 2 件一级品，7 件二级品和 1 件次品. 从 10 件产品中不放回地抽取了 3 件. 用 ξ 表示其中的一级品数，η 表示其中的二级品数. 求 ξ 与 η 的相关系数.

33. 在 polya 罐子模型(罐子中有 b 个黑球，r 个红球，随机地取出一个球，然后再放回罐中，并且还多放 c 个同色的球)中，设 ξ_n 是 1 或 0，按第 n 次试验的结果是红球或黑球而定. 证明：对 $m\neq n$，$\rho_{\xi_m,\xi_n}=c/b+r+c$.

---

① 游程是概率论中一个专门术语. 在掷硬币的贝努里试验序列中，若连续出现同一种结果 k 次，就称这结果的游程长度为 k. 这样，若用 S 表示出现正面，F 表示出现背面，则由第一个试验开始的游程(称为第一个游程，紧跟着的下一个游程称为第二个游程)，长度为 2 是结果为 $SSF\cdots$ 或 $FFS\cdots$ 的所有序列.

34. 设随机变数(ξ,η)的密度函数为

$$f(x,y)=A\sin(x+y),$$

$$0\leqslant x\leqslant\frac{\pi}{2},\quad 0\leqslant y\leqslant\frac{\pi}{2}.$$

求：i) 系数A；ii) $E(\xi)$、$E(\eta)$、$D(\xi)$、$D(\eta)$；iii) 协方差及相关系数ρ.

35. 设随机变数$\xi_1,\xi_2,\cdots,\xi_{n+m}(n>m)$独立同分布且有有限方差. 试求$\eta=\sum_{k=1}^{n}\xi_k$与$\zeta=\sum_{k=1}^{n}\xi_{m+k}$之间的相关系数.

36. 证明：对随机变数ξ、η，$E(\xi\eta)=E(\xi)\cdot E(\eta)$或$D(\xi+\eta)=D(\xi)+D(\eta)$的充要条件为：$\rho=0$.

37. 试证明，如果随机变数ξ与η都取两个值，且协方差为零，则ξ与η相互独立.

38. 证明：如果随机变数ξ,η相互独立，则

$$D(\xi\eta)=D(\xi)\cdot D(\eta)+[E(\xi)]^2D(\eta)+[E(\eta)]^2D(\xi).$$

39. 设$x_1,x_2,\cdots,x_k$是随机变数ξ的可能值并且为正值. 求证：当$n\to\infty$时，

a) $\dfrac{E(\xi^{n+1})}{E(\xi^n)}\to\max\limits_{1\leqslant j\leqslant k}x_j$;

b) $\sqrt[n]{E(\xi^n)}\to\max\limits_{1\leqslant j\leqslant k}x_j$.

40. 设随机变数ξ的分布函数为$F(x)$，求证：如果$E(\xi)$存在，则

$$E(\xi)=\int_0^{\infty}[1-F(x)+F(-x)]dx.$$

并且，要使$E(\xi)$存在，必须有

$$\lim_{x\to-\infty}xF(x)=\lim_{x\to\infty}x[1-F(x)]=0.$$

41. 已知(ξ,η)的联合密度函数为

$$f(x,y)=\begin{cases}3x, & 0<y<x,\ 0<x<1,\\ 0, & \text{其它}.\end{cases}$$

求$E\left(\xi\mid\eta=\frac{1}{2}\right)$及$D\left(\xi\mid\eta=\frac{1}{2}\right)$.

42. 设每天到达货站的货物的件数N的概率分布是：

$N=n$	10	11	12	13	14	15
$P(N=n)$	0.05	0.10	0.10	0.20	0.35	0.20

如果每天到达的货物中次品的概率是相同的，都等于 0.10. 用 ξ 表示每天到达的货物中次品的件数，求 $E\xi$.

第四章 特征函数

前面讲过，我们通过随机变数来描述随机现象．通过随机变数的分布函数来了解它的统计规律，通过数字特征来掌握分布函数的某些特征．我们看到，数字特征一般由各阶矩所决定，但随着阶数的增高，求矩的计算会愈加麻烦．另一方面，随机现象是错综复杂的．往往需要多个随机变数，甚至要由一随机变数序列依某种收敛意义的逼近来描述．在第二章里，我们已经看到，即使是两个相互独立的随机变数之和或积，求它们的分布函数也不是很简单的．总之，要解决复杂得多的问题没有优越的数学工具是不行的．大家知道，富里埃变换是数学中非常重要而有效的工具，把它应用于分布函数或密度函数，就产生了所谓"特征函数"．而当随机变数为离散型时，则用"母函数"更为方便，因为这时可以充分利用幂级数的性质而避免引进复函数的积分．

可以毫不夸张地说，概率论和数理统计自从引进了母函数，特别是特征函数以后，就把理论的研究推进到一个新的阶段．

§4.1 特征函数的定义及其性质

一、特征函数定义及例

定义 4.1.1 设 ξ 是定义在概率空间 $(\Omega, \mathscr{F}, P)$ 上的随机变数，它的分布函数为 $F(x)$，称 $e^{jt\xi}$ 的数学期望 $E(e^{jt\xi})$ 为 ξ 的特征函数，有时也称为分布函数 $F(x)$ 的特征函数，其中 $j=\sqrt{-1}$，$t\in R_1$．记 ξ 的特征函数为 $\varphi_\xi(t)$，在不引起混乱的情况下简写为 $\varphi(t)$．

对复的随机变数的数学期望定义如下：

若复随机变数为 $z=X+jY$，其中 X,Y 均为实随机变数，则 z 的数学期望定义为

$$E(z)=E(X)+jE(Y).$$

读者可以证明，在第三章中所述关于实随机变数数学期望的性质，对复随机变数来说也成立. 由于

$$e^{jt\xi}=\cos t\xi+j\sin t\xi,$$

故

$$\begin{aligned}\varphi(t)=E(e^{jt\xi})&=E(\cos t\xi)+jE(\sin t\xi)\\&=\int_{-\infty}^{\infty}\cos tx dF(x)+j\int_{-\infty}^{\infty}\sin tx dF(x)\\&=\int_{-\infty}^{\infty}e^{jtx}dF(x).\end{aligned}\tag{4.1.1}$$

因此，ξ 的特征函数也可称之为对分布函数 $F(x)$ 的富里埃-司蒂阶变换.

因为对任意 $t\in R_1$，$\cos tx$ 及 $\sin tx$ 皆为有界连续函数，故 $E(\cos t\xi)$ 及 $E(\sin t\xi)$ 皆为有限数. 因此，任一随机变数的特征函数总是存在的.

由定义 4.1.1 及(3.4.2)，(3.4.3) 两式得到：当 ξ 为离散型随机变数时，其特征函数为

$$\varphi(t)=E(e^{jt\xi})=\sum_k e^{jtx_k}p_k,\tag{4.1.2}$$

其中 $p_k=P\{\xi=x_k\}$.

当 ξ 为连续型随机变数时，其特征函数为

$$\varphi(t)=E(e^{jt\xi})=\int_{-\infty}^{\infty}e^{jtx}f(x)dx.\tag{4.1.3}$$

由(4.1.2)及(4.1.3)两式看到，要计算特征函数时，需要进行复数运算(复数求和)或进行实变复值函数的积分. 对前一种情况可直接按复数运算法则进行计算或把复数的实部和虚部分别按一

般实数求和进行计算，在后一种复函数积分的情况，有时则要应用复函数的留数理论.

下面例 4.1.1——4.1.4 都可以按复数运算直接计算. 我们知道对复数的二项式展开总是成立的，即

$$(z_1+z_2)^n=\sum_{k=0}^{n}C_n^k z_1^k z_2^{n-k}, \tag{4.1.4}$$

其中 z_1, z_2 为复数. 且 e^z 有如下表达式：对任意复数 z

$$e^z=1+\frac{z}{1!}+\frac{z^2}{2!}+\cdots+\frac{z^n}{n!}+\cdots, \tag{4.1.5}$$

而且(4.1.5)对任意复数 z 都绝对收敛.

例 4.1.1 设随机变数 ξ 服从退化分布，即

$$P\{\xi=c\}=1,$$

其中 c 为某一常数. 则有

$$\varphi(t)=E(e^{jtc})=e^{jtc}.$$

例 4.1.2 设随机变数 ξ 服从两点分布，即

$$P\{\xi=1\}=p,\ P\{\xi=0\}=1-p=q.$$

则其特征函数

$$\varphi(t)=e^{jt\cdot 0}q+e^{jt\cdot 1}p=q+pe^{jt}.$$

例 4.1.3 设随机变数 ξ 服从二项分布，即

$$p_k=P\{\xi=k\}=C_n^k p^k q^{n-k},\quad (k=0,1,\cdots,n).$$

则有

$$\varphi(t)=\sum_{k=0}^{n}e^{jtk}p_k=\sum_{k=0}^{n}C_n^k p^k q^{n-k}e^{jtk}$$

$$=\sum_{k=0}^{n}C_n^k(pe^{jt})^k q^{n-k}=(q+pe^{jt})^n$$

例 4.1.4 设随机变数 ξ 服从泊松分布，即

$$p(k)=P\{\xi=k\}=\frac{\lambda^k}{k!}e^{-\lambda},\ \text{常数}\ \lambda>0,$$

$$(k=0,1,\cdots).$$

则

$$\varphi(t)=\sum_{k=0}^{\infty}e^{jtk}p_k=\sum_{k=0}^{\infty}e^{jtk}\frac{\lambda^k}{k!}e^{-\lambda}$$

$$=e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{jt})^k}{k!}=e^{-\lambda}e^{\lambda e^{jt}}=e^{\lambda(e^{jt}-1)}.$$

例 4.1.5 设随机变量 ξ 服从 $[-a,a]$ 上的均匀分布，即 ξ 的密度函数为

$$f(x)=\begin{cases}\frac{1}{2a}, & x\in[-a,a];\\ 0, & \text{其他},\end{cases}$$

则

$$\varphi(t)=\int_{-a}^{a}e^{jtx}\cdot\frac{1}{2a}dx=\frac{1}{2a}\left[\int_{-a}^{a}\cos tx dx+j\int_{-a}^{a}\sin tx dx\right]$$

$$=\frac{1}{2a}\cdot 2\frac{\sin at}{t}=\frac{\sin at}{at}.$$

例 4.1.6 设随机变量 ξ 服从指数分布，即

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\geqslant 0;\\ 0 & \text{其他},\end{cases}$$

其中常数 $\lambda>0$，则

$$\varphi(t)=\int_0^{\infty}\lambda e^{jtx}e^{-\lambda x}dx=\int_0^{\infty}\lambda e^{-\lambda x}\cos tx dx+j\lambda\int_0^{\infty}e^{-\lambda x}\sin tx dx$$

$$=\lambda\frac{\lambda}{\lambda^2+t^2}+j\lambda\frac{t}{\lambda^2+t^2}=\left(1-j\frac{t}{\lambda}\right)^{-1}.$$

例 4.1.7 设随机变量 ξ 服从正态 $N(0,1)$，则其特征函数

$$\varphi(t)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{jtx}e^{-\frac{1}{2}x^2}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{jtx}e^{-\frac{1}{2}x^2}dx$$

由于

$$|jxe^{jtx-\frac{1}{2}x^2}|\leqslant|x|e^{-\frac{1}{2}x^2},$$

且

$$\int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}x^2} dx < \infty.$$

故有

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} jx e^{jtx-\frac{1}{2}x^2} dx.$$

易见

$$jt\varphi(t) + j\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (jt-x) e^{jtx-\frac{1}{2}x^2} dx.$$

$$= \frac{1}{\sqrt{2\pi}} \left[e^{jtx-\frac{1}{2}x^2} \right]_{-\infty}^{\infty} = 0.$$

故得微分方程

$$t\varphi(t) + \varphi'(t) = 0.$$

解微分方程

$$\frac{\varphi'(t)}{\varphi(t)} = -t$$

得

$$\ln\varphi(t) = -\frac{1}{2}t^2 + C,$$

即

$$\varphi(t) = e^C e^{-\frac{1}{2}t^2}.$$

令 $t=0$，得

$$1 = \varphi(0) = e^C, \quad \therefore \quad C = 0.$$

故求得

$$\varphi(t) = e^{-\frac{1}{2}t^2}.$$

用同样的方法可以证明随机变数 ξ 服从正态 $N(a,\sigma)$ 时，则其特征函数为：

$$\varphi(t) = e^{jat-\frac{1}{2}\sigma^2 t^2}.$$

例 4.1.8 设随机变数 ξ 在区间 $[-2, 2]$ 上服从三角形分布，即 ξ 有密度函数

$$f(x)=\begin{cases}\frac{1}{4}(2+x) & \text{当} -2\leqslant x<0;\\ \frac{1}{4}(2-x) & \text{当}\ 0\leqslant x\leqslant 2;\\ 0 & \text{当}\ |x|>2,\end{cases}$$

则其特征函数为

$$\varphi(t)=\int_{-2}^{0}\frac{1}{4}(2+x)e^{jtx}dx+\int_{0}^{2}\frac{1}{4}(2-x)e^{jtx}dx.$$

而

$$\int_{-2}^{0}(2+x)e^{jtx}dx=\frac{1}{jt}(2+x)e^{jtx}\Big|_{-2}^{0}-\frac{1}{jt}\int_{-2}^{0}e^{jtx}dx$$

$$=\frac{2}{jt}-\frac{1}{jt}\cdot\frac{1}{jt}e^{jtx}\Big|_{-2}^{0}$$

$$=\frac{2}{jt}+\frac{1}{t^2}[1-e^{-2jt}].$$

同理

$$\int_{0}^{2}(2-x)e^{jtx}\,dx=-\frac{2}{jt}+\frac{1}{t^2}[1-e^{2jt}].$$

因而

$$\varphi(t)=\frac{1}{4}\left[\frac{2-e^{2jt}-e^{-2jt}}{t^2}\right]=\frac{1}{2t^2}(1-\cos 2t)$$

$$=\left(\frac{\sin t}{t}\right)^2.$$

例 4.1.9 设随机变数 ξ 服从哥西分布，即 ξ 的密度函数为

$$f(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2},$$

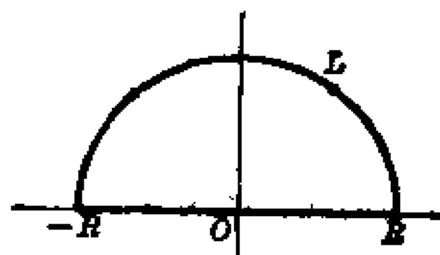


图 4.1.1

其中 $\lambda>0$ 及 μ 均为常数. 则其特征函数为

$$\varphi(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx.$$

考虑积分路线 L 如图 4.1.1 中箭头指示方向的复变积分

$$\int_L e^{jtz}\frac{\lambda}{\lambda^2+(z-u)^2}dz.$$

由复变函数留数理论可知当 $t>0$ 时

$$\int_{-R}^{R}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx+\int_{C_R}e^{jtz}\frac{\lambda}{\lambda^2+(z-\mu)^2}dz$$

$$=\int_L e^{jtz}\frac{\lambda}{\lambda^2+(z-\mu)^2}dz$$

$$=2\pi j\lim_{z\to\mu+j\lambda}(z-\mu-j\lambda)e^{jtz}\frac{\lambda}{\lambda^2+(z-\mu)^2}$$

$$=2\pi je^{jt(\mu+j\lambda)}\frac{\lambda}{2j\lambda}=\pi e^{jt\mu}e^{-t\lambda}.$$

其中 C_R 表示积分路线 L 中的半圆周. 而由约当引理可知

$$\int_{C_R}e^{jtz}\frac{\lambda}{\lambda^2+(z-\mu)^2}dz\to 0\quad(\text{当 }R\to\infty\text{ 时}).$$

由此得

$$\int_{-\infty}^{\infty}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx=\pi e^{jt\mu}e^{-t\lambda},$$

同理可得 $t<0$ 时

$$\int_{-\infty}^{\infty}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx=\pi e^{jt\mu}e^{t\lambda}.$$

当 $t=0$ 时显然有

$$\int_{-\infty}^{\infty}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx=\pi.$$

综上所述我们得到

$$\varphi(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}e^{jtx}\frac{\lambda}{\lambda^2+(x-\mu)^2}dx=e^{jt\mu}e^{-|t|\lambda}.$$

特别，例 4.1.9 中当 $\mu=0,\lambda=1$ 时，即 ξ 服从具有密度函数

$$f(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$$

的哥西分布时，其特征函数为

$$\varphi(t)=e^{-|t|}.$$

二、特征函数性质

性质 4.1.1 随机变数 ξ 的特征函数 $\varphi(t)$ 满足：

(1) $|\varphi(t)|\leq\varphi(0)=1$;

(2) $\varphi(-t)=\overline{\varphi(t)}$，($\overline{\varphi(t)}$ 表示 $\varphi(t)$ 的共轭复数)。

证(1) ∵ $\varphi(t)=E(e^{jt\xi})=E(\cos t\xi)+jE(\sin t\xi)$,

$$\begin{aligned}\therefore|\varphi(t)|^2&=E^2(\cos t\xi)+E^2(\sin t\xi)\\&\leq E(\cos t\xi)^2+E(\sin t\xi)^2\\&=E(\cos^2 t\xi+\sin^2 t\xi)=1.\end{aligned}$$

另外

$$\varphi(0)=\int_{-\infty}^{\infty}dF(x)=1,$$

故有

$$|\varphi(t)|\leq\varphi(0)=1.$$

现证(2)，∵ $\varphi(-t)=E(e^{-jt\xi})$

$$\begin{aligned}&=E[\cos(-t\xi)]+jE[\sin(-t\xi)]\\&=E[\cos t\xi]-jE[\sin t\xi]=\overline{\varphi(t)}.\ ∎\end{aligned}$$

性质 4.1.2 设 ξ 的特征函数为 $\varphi_\xi(t)$，则 $\eta=a\xi+b$ 的特征函数 $\varphi_\eta(t)$ 为

$$\varphi_\eta(t)=e^{jbt}\varphi_\xi(at)\text{，其中 } a,b \text{ 为常数}$$

证 $\varphi_\eta(t)=E(e^{j\eta t})=E[e^{j(a\xi+b)t}]$

$$=E(e^{jbt}\cdot e^{ja\xi t})=e^{jbt}\varphi_\xi(at).\ ∎$$

例 4.1.10 设随机变数 ξ 服从 $N(0,1)$，则

$$\eta=\sigma\xi+a\ (\text{其中 } \sigma>0 \text{ 及 } a \text{ 均为常数})$$

的特征函数为

$$\varphi_\eta(t)=e^{jat}\varphi_\xi(\sigma t).$$

由例 4.1.7 知，若 ξ 服从正态 $N(0,1)$，则

$$\varphi_\xi(\sigma t)=e^{-\frac{1}{2}(\sigma t)^2}.$$

故 $\eta=\sigma\xi+a$ 的特征函数为

$$\varphi_\eta(t)=e^{jat}e^{-\frac{1}{2}\sigma^2t^2}.$$

即此特征函数与服从 $N(a,\sigma)$ 的随机变数的特征函数一样.

性质 4.1.3 随机变数 ξ 的特征函数 $\varphi(t)$ 在 R_1 上一致连续.

证 $\because$

$$\begin{aligned}|\varphi(t+h)-\varphi(t)|&=|E(e^{j(t+h)\xi})-E(e^{jt\xi})|\\&=|E[e^{j(t+h)\xi}-e^{jt\xi}]|.\end{aligned}$$

而

$$e^{j(t+h)\xi}-e^{jt\xi}=e^{jt\xi}(e^{jh\xi}-1),$$

故 $\quad |e^{j(t+h)\xi}-e^{jt\xi}|=|e^{jh\xi}-1|=|e^{j\frac{h}{2}\xi}(e^{j\frac{h}{2}\xi}-e^{-j\frac{h}{2}\xi})|.$

这样一来得

$$|\varphi(t+h)-\varphi(t)|\leqslant\int_{-l}^{l}2\left|\sin\frac{hx}{2}\right|dF(x)+2\int_{|x|\geqslant l}dF(x).$$

对任给 $\varepsilon>0$，取充分大的 l，使得

$$\int_{|x|\geqslant l}dF(x)<\frac{\varepsilon}{4},$$

再取 $\delta>0$，当 $|h|<\delta$ 时，对一切 $x\in[-l,l]$ 一致地有

$$\left|\sin\frac{h}{2}x\right|<\frac{\varepsilon}{4},$$

因此最后得到，对 t 一致地有

$$\begin{aligned}|\varphi(t+h)-\varphi(t)|&<\frac{\varepsilon}{2}\int_{-\infty}^{\infty}dF(x)+\frac{\varepsilon}{2}\\&=\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\ \blacksquare\end{aligned}$$

性质 4.1.4 随机变数 ξ 的特征函数 $\varphi(t)$ 是非负定的，即对任意正整数 n，任意复数 $z_1,z_2,\cdots,z_n$ 以及 $t_r\in R_1$，$r=1,2,\cdots,n$，有

$$\sum_{r,s=1}^{n}\varphi(t_r-t_s)z_r\bar{z}_s\geqslant 0.$$

证 ∵ $$\sum_{r,s=1}^{n}\varphi(t_r-t_s)z_r\bar{z}_s=\sum_{r,s=1}^{n}z_r\bar{z}_s\int_{-\infty}^{\infty}e^{j(t_r-t_s)x}dF(x)$$

$$=\int_{-\infty}^{\infty}\left[\sum_{r,s=1}^{n}z_r\bar{z}_s e^{jt_rx}e^{-jt_sx}\right]dF(x)$$

$$=\int_{-\infty}^{\infty}\left|\sum_{r=1}^{n}z_r e^{jt_rx}\right|^2dF(x)\geqslant 0.\ \blacksquare$$

上面我们证明了，特征函数 $\varphi(t)$ 一致连续、非负定且 $\varphi(0)=1$. 可以证明逆命题也是成立的，即有波赫纳-辛钦定理：

若函数 $\varphi(t)$, $(t\in R_1)$ 连续、非负定且 $\varphi(0)=1$，则 $\varphi(t)$ 必为特征函数（证明参阅[1]或[2]）.

三、特征函数与矩的关系

在第三章里，我们看到求随机变数 ξ 的各阶矩，有时需要作非常繁杂的积分计算，有了特征函数后，求矩的方法就可通过对特征函数求导数的办法来得到. 关于这一点有如下的定理：

定理 4.1.1 设随机变数 ξ 的 n 阶矩存在，则 ξ 的特征函数 $\varphi(t)$ 的 k 阶导数 $\varphi^{(k)}(t)$ 存在，且

$$E(\xi^k)=j^{-k}\varphi^{(k)}(0)\qquad(k\leqslant n).\tag{4.1.6}$$

证 现只对 ξ 是连续型的情况加以证明：

设 $f(x)$ 为 ξ 的密度函数，在(4.1.3)式中的被积函数 $e^{jtx}f(x)$ 对 t 的 k 阶导数为

$$j^kx^ke^{jtx}f(x),$$

而

$$\int_{-\infty}^{\infty}|j^kx^ke^{jtx}f(x)|dx=\int_{-\infty}^{\infty}|x^kf(x)|dx=\beta_k<\infty.$$

由此可知(4.1.3)式右端可以在积分号下求 k 阶 $(k\leqslant n)$ 导数，因此

$$\varphi^{(k)}(t)=\int_{-\infty}^{\infty} j^k x^k e^{jtx} f(x)dx = j^k \int_{-\infty}^{\infty} x^k e^{jtx} f(x)dx$$
$$= j^k E(\xi^k e^{jt\xi}).$$

令 $t=0$，得

$$\varphi^{(k)}(0)=j^k E(\xi^k),$$

因此得

$$E(\xi^k)=j^{-k}\varphi^{(k)}(0). \quad ▌ \tag{4.1.7}$$

例 4.1.11 设随机变数 ξ 服从泊松分布，由例 4.1.4 已知它的特征函数为

$$\varphi(t)=e^{\lambda(e^{jt}-1)},$$

故

$$\varphi'(t)=j\lambda e^{jt}e^{\lambda(e^{jt}-1)},$$
$$\varphi'(0)=j\lambda.$$

按(4.1.7)式得 ξ 的数学期望

$$E(\xi)=(-1)j\varphi'(0)=\lambda.$$

又

$$\varphi''(t)=\lambda j^2 e^{jt} e^{\lambda(e^{jt}-1)}(\lambda e^{jt}+1),$$
$$\varphi''(0)=-\lambda(\lambda+1),$$

由此得

$$E(\xi^2)=(-1)^2 j^2 \varphi''(0)=-[-\lambda(\lambda+1)]=\lambda(\lambda+1).$$

这样 ξ 的方差为

$$D(\xi)=E(\xi^2)-[E(\xi)]^2=\lambda(\lambda+1)-\lambda^2=\lambda.$$

用相同的办法可求得更高阶的矩.

例 4.1.12 设随机变数 ξ 服从 $N(a,\sigma)$，由例 4.1.7 已知 ξ 的特征函数为

$$\varphi(t)=e^{jat-\frac{1}{2}\sigma^2t^2}.$$

故

$$\varphi'(t)=(ja-\sigma^2 t)e^{jat-\frac{1}{2}\sigma^2t^2},$$

$$\varphi''(t)=[(ja-\sigma^2 t)^2-\sigma^2]e^{jat-\frac{1}{2}\sigma^2 t^2},$$

$$\varphi'(0)=ja;\ \ \varphi''(0)=(ja)^2-\sigma^2=-a^2-\sigma^2.$$

由此得

$$E(\xi)=-j(ja)=a.$$

$$E(\xi^2)=(-1)^2 j^2\varphi''(0)=-(-a^2-\sigma^2)=a^2+\sigma^2.$$

这样，ξ 的方差为

$$D(\xi)=E(\xi^2)-[E(\xi)]^2=a^2+\sigma^2-a^2=\sigma^2.$$

上述所求的数学期望与方差，我们在第三章已经计算过.

对于例 4.1.12，读者还可以进一步计算其 k 阶中心矩为

$$\mu_k=E(\xi-a)^k=\begin{cases}0, & k\text{ 为奇数};\\ 1\cdot 3\cdot 5\cdot\cdots\cdot(k-1)\sigma^k, & k\text{ 为偶数}.\end{cases}$$

§ 4.2 反演公式及唯一性定理

对于给定的一个分布函数 $F(x)$，公式(4.1.1)唯一决定它的特征函数 $\varphi(t)$. 下面定理说明其逆也真，即分布函数 $F(x)$能唯一地被其特征函数 $\varphi(t)$表达出来. 这就建立了分布函数与特征函数一一对应的关系.

定理 4.2.1（反演公式） 设随机变数 ξ 的分布函数和特征函数分别为 $F(x)$ 和 $\varphi(t)$，则对于 $F(x)$ 的任意连续点 x_1 和 x_2 $(-\infty<x_1<x_2<\infty)$，有

$$F(x_2)-F(x_1)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-jtx_1}-e^{-jtx_2}}{jt}\varphi(t)\,dt\text{①}. \quad (4.2.1)$$

令

① 如果 x_1, x_2 不是 $F(x)$ 的连续点，则可证明

$$\frac{F(x_2+0)+F(x_2)}{2}-\frac{F(x_1+0)+F(x_1)}{2}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-jtx_1}-e^{-jtx_2}}{jt}\varphi(t)\,dt. \quad (4.2.1)'$$

$a=\frac{x_1+x_2}{2}$, $h=\frac{x_2-x_1}{2}>0$,则可以把 x_1, x_2 写成

$$x_1=a-h,\quad x_2=a+h$$

这样(4.2.1)就可以写成下面完全等价的形式:

$$F(a+h)-F(a-h)=\lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sin th}{t}e^{-jta}\varphi(t)\,dt. \tag{4.2.2}$$

证 我们只对具有密度函数 $f(x)$ 的连续型随机变数来证明(4.2.1),为此,先证如下引理:

引理 4.2.1 对任意实数 α 及 $T>0$,令

$$I(\alpha,T)=\frac{2}{\pi}\int_0^T\frac{\sin\alpha t}{t}dt, \tag{4.2.3}$$

则有

$$\lim_{T\to\infty}I(\alpha,T)=\begin{cases}1, & 当\ \alpha>0;\\ 0, & 当\ \alpha=0;\\ -1, & 当\ \alpha<0,\end{cases} \tag{4.2.4}$$

并且,对 $|\alpha|>\delta$ 一致收敛,其中 δ 为任意小的正数。

证 $\because\quad \lim\limits_{T\to\infty}\int_0^T\frac{\sin t}{t}dt=\frac{\pi}{2}$[①],

而

$$I(\alpha,T)=\frac{2}{\pi}\int_0^T\frac{\sin\alpha t}{t}dt=\frac{2}{\pi}\int_0^{\alpha T}\frac{\sin t}{t}dt.$$

故得

$$\lim_{T\to\infty}I(\alpha,T)=\lim_{T\to\infty}\frac{2}{\pi}\int_0^{\alpha T}\frac{\sin t}{t}dt=\begin{cases}1 & \alpha>0;\\ 0 & \alpha=0;\\ -1 & \alpha<0.\end{cases}$$

又由于

$$\lim_{T\to\infty}\int_0^T\frac{\sin t}{t}dt=\frac{\pi}{2},$$

---

① 参看江泽坚著"数学分析"下册第十三章§4或其他微积分教本.

故对任意 $\varepsilon>0$,存在 T_1,使当 $T>T_1$ 时

$$\left|\int_0^T \frac{\sin t}{t} dt-\frac{\pi}{2}\right|<\varepsilon \quad 或 \quad \left|\frac{2}{\pi}\int_0^T \frac{\sin t}{t} dt-1\right|<\varepsilon,$$

若 $\alpha>\delta>0$, 则当 $T>T_2=\frac{T_1}{\delta}$时,对一切 $\alpha>\delta>0$,有

$$|I(\alpha,T)-1|<\varepsilon.$$

同理可证 $\alpha<-\delta<0$ 时的情形. 因而我们证明了当 $|\alpha|>\delta$ 时(4.2.4)式的一致收敛性. ▌

引理 4.2.2 令

$$I(T,x,a,b)=\frac{1}{2\pi}\int_{-T}^{T}\frac{\sin t(x-a)-\sin t(x-b)}{t}dt,$$

则

$$\lim_{T\to\infty}I(T,x,a,\ b)=\begin{cases}1, & 当\ a<x<b;\\ \frac{1}{2}, & 当\ x=a\ 或\ x=b;\\ 0, & 当\ x<a\ 或\ x>b,\end{cases}$$

并且对 $|x-a|>\delta$ 及 $|x-b|>\delta$ 是一致收敛的. 其中 δ 为任意小的正数.

证 显然

$$I(T,x,a,b)=\frac{1}{2}[I(x-a,T)-I(x-b,T)],$$

其中 $I(x,T)$ 由(4.2.3)式所定义.

由引理 4.2.1 立即得到

$$\lim_{T\to\infty}I(T,x,a,\ b)=\begin{cases}1, & 当\ a<x<b;\\ \frac{1}{2}, & 当\ x=a\ 或\ x=b;\\ 0, & 当\ x<a\ 或\ x>b,\end{cases}$$

以及对 $|x-a|>\delta$ 及 $|x-b|>\delta$,上式一致收敛. ▌

下面证明定理 4.2.1:

令

$$\phi(t)=\frac{e^{-jtx_1}-e^{-jtx_2}}{jt}\varphi(t),$$

$$J(T)=\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-jtx_1}-e^{-jtx_2}}{jt}\varphi(t)dt,$$

因为

$$\varphi(-t)=\overline{\varphi(t)},$$

得

$$\phi(-t)=\overline{\phi(t)}.$$

用 $\mathrm{Im}\{\phi(t)\}$ 及 $\mathrm{Re}\{\phi(t)\}$ 表示 $\phi(t)$ 的虚部和实部，则对任意 $T>0$，有

$$\int_{-T}^{T}\mathrm{Im}\{\phi(t)\}dt=0.$$

且

$$\begin{aligned}\mathrm{Re}\{\phi(t)\}&=\mathrm{Re}\left\{\int_{-\infty}^{\infty}\frac{e^{-jtx_1}-e^{-jtx_2}}{jt}e^{jtx}f(x)dx\right\}\\&=\mathrm{Re}\left\{\int_{-\infty}^{\infty}\frac{e^{jt(x-x_1)}-e^{jt(x-x_2)}}{jt}f(x)dx\right\}\\&=\int_{-\infty}^{\infty}\frac{\sin t(x-x_1)-\sin t(x-x_2)}{t}f(x)dx\end{aligned}$$

这样

$$\begin{aligned}J(T)&=\frac{1}{2\pi}\left\{\int_{-T}^{T}\mathrm{Re}\{\phi(t)\}dt+j\int_{-T}^{T}\mathrm{Im}\{\phi(t)\}dt\right\}\\&=\frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{\infty}\frac{\sin t(x-x_1)-\sin t(x-x_2)}{t}f(x)dxdt.\end{aligned}$$

上式的积分次序可以交换，这是因为对 t 的积分限是有限的，而对 x 的积分关于 t 一致收敛，后者可由下式看出：

$$\left|\frac{\sin t(x-x_1)-\sin t(x-x_2)}{t}f(x)\right|$$

$$= \left| \frac{2\cos t\frac{2x-x_2-x_1}{2}\sin t\frac{x_2-x_1}{2}}{t} \right| f(x)$$

$$\leqslant 2\left| \frac{\sin t\frac{x_2-x_1}{2}}{t} \right| f(x) = 2\cdot\frac{x_2-x_1}{2}\cdot\left| \frac{\sin t\frac{x_2-x_1}{2}}{t\frac{x_2-x_1}{2}} \right| f(x)$$

$$\leqslant (x_2-x_1)f(x).$$

其中 $f(x)$为密度函数，有

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

因此

$$J(T) = \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-T}^{T}\frac{\sin t(x-x_1)-\sin t(x-x_2)}{t}dt\right]f(x)dx$$

$$= \int_{-\infty}^{\infty} I(T,x,x_1,x_2)f(x)dx$$

$$= \int_{-\infty}^{x_1-\delta} I(T,x,x_1,x_2)f(x)dx$$

$$+ \int_{x_1-\delta}^{x_1+\delta} I(T,x,x_1,x_2)f(x)dx$$

$$+ \int_{x_1+\delta}^{x_2-\delta} I(T,x,x_1,x_2)f(x)dx$$

$$+ \int_{x_2-\delta}^{x_2+\delta} I(T,x,x_1,x_2)f(x)dx$$

$$+ \int_{x_2+\delta}^{\infty} I(T,x,x_1,x_2)f(x)dx.$$

由引理 4.2.2 得

$$\lim_{T\to\infty}\int_{-\infty}^{x_1-\delta} I(T,x,x_1,x_2)f(x)dx$$

$$= \int_{-\infty}^{x_1-\delta}\lim_{T\to\infty} I(T,x,x_1,x_2)f(x)dx = 0.$$

同理

$$\lim_{T\to\infty}\int_{x_2+\delta}^{\infty} I(T,x,x_1,x_2)f(x)dx=0.$$

又

$$\lim_{T\to\infty}\int_{x_1+\delta}^{x_2-\delta} I(T,x,x_1,x_2)f(x)dx$$

$$=\int_{x_1+\delta}^{x_2-\delta}\lim_{T\to\infty} I(T,x,x_1,x_2)f(x)dx$$

$$=\int_{x_1+\delta}^{x_2-\delta} f(x)dx=F(x_2-\delta)-F(x_1+\delta), \tag{4.2.5}$$

而在积分

$$\int_{x_1-\delta}^{x_1+\delta} I(T,x,x_1,x_2)f(x)dx$$

中，被积函数有

$$|I(T,x,x_1,x_2)|\leqslant M,$$

故

$$\left|\int_{x_1-\delta}^{x_1+\delta} I(T,x,x_1,x_2)f(x)dx\right|\leqslant M\int_{x_1-\delta}^{x_1+\delta} f(x)dx$$

$$=M[F(x_1+\delta)-F(x_1-\delta)]. \tag{4.2.6}$$

同理

$$\left|\int_{x_2-\delta}^{x_2+\delta} I(T,x,x_1,x_2)f(x)dx\right|\leqslant M[F(x_2+\delta)-F(x_2-\delta)]. \tag{4.2.7}$$

由于 x_1,x_2 为 $F(x)$ 的连续点，故当 $\delta\to0$ 时，(4.2.5)式右端等于 $F(x_2)-F(x_1)$，同时(4.2.6)及(4.2.7)两式都趋于零，因此就有

$$\lim_{T\to\infty} J(T)=F(x_2)-F(x_1). \quad \blacksquare$$

定理 4.2.1 说明，当 x_1,x_2 为 $F(x)$ 的连续点时，$F(x_2)-F(x_1)$ 的值完全由特征函数按(4.2.1)式给出，即它给出了随机变数 ξ 取值于区间 $[x_1,x_2)$ 的概率：

$$P\{x_1\leqslant\xi<x_2\}=F(x_2)-F(x_1).$$

设 $x_2=x$ 为 $F(x)$ 的连续点，令 x_1 沿着 $F(x)$ 的连续点趋于 $-\infty$，这时

$$\lim_{x_1\to-\infty} F(x_1)=0,$$

由此得到

$$F(x)=\lim_{x_1\to-\infty}[F(x)-F(x_1)]$$

$$=\lim_{x_1\to-\infty}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-jtx_1}-e^{-jtx}}{jt}\varphi(t)dt. \qquad (4.2.8)$$

这样，对于 $F(x)$的每一连续点，$F(x)$由特征函数按(4.2.8)式完全给定. 如果对于 $F(x)$的不连续点 x，我们重新规定

$$\bar{F}(x)=\frac{F(x+0)+F(x)}{2} \qquad (4.2.9)$$

则对任意 $x\in R_1$，$\bar{F}(x)$完全由特征函数 $\varphi(t)$按(4.2.1)式给出.

特征函数根据(4.2.1)可确定分布函数. 我们把公式(4.2.1)叫做反演公式.

推论 1(唯一性定理) 分布函数 $F_1(x)$ 及 $F_2(x)$ 恒等的充分必要条件为它们的特征函数 $\varphi_1(t)$ 及 $\varphi_2(t)$ 恒等.

证 由 (4.1.1) 式，当 $F_1(x)=F_2(x)$时，显然对一切 $t\in R_1$ $\varphi_1(t)=\varphi_2(t)$. 反之，若 $\varphi_1(t)=\varphi_2(t)$，则由式(4.2.8)得到，对一切是 $F_1(x)$且也是 $F_2(x)$ 的连续点 x (这些点的全体记为 A)，有 $F_1(x)=F_2(x)$；对于不是 $F_1(x)$或 $F_2(x)$的连续点 x，总可以取一列单调上升的 $x_n\in A$，$x_n\to x$，并根据分布函数的左连续性，使得

$$F_1(x)=\lim_{n\to\infty}F_1(x_n)=\lim_{n\to\infty}F_2(x_n)=F_2(x),$$

这样就得到了对一切 $x\in R_1$，有

$$F_1(x)=F_2(x). ∎$$

推论 2 若随机变数 ξ 的特征函数 $\varphi(t)$于 R_1 上绝对可积，则 ξ 为具有密度函数 $f(x)$的连续型随机变数，且

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-jtx}\varphi(t)dt. \tag{4.2.10}$$

证 由(4.2.9)及(4.2.1)′两式，令 $x_2=x+\Delta x$(此处 $\Delta x>0$)，$x_1=x$，则有

$$\begin{aligned}
F(x+\Delta x)-F(x)&=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-jtx}-e^{-jt(x+\Delta x)}}{jt}\varphi(t)dt\\
&=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{1-e^{-jt\Delta x}}{jt}e^{-jtx}\varphi(t)dt\\
&=\frac{1}{\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{\sin t\frac{\Delta x}{2}}{t}\cdot e^{-jt\left(x+\frac{\Delta x}{2}\right)}\varphi(t)dt.
\end{aligned}$$

故

$$\frac{F(x+\Delta x)-F(x)}{\Delta x}=\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{\sin t\frac{\Delta x}{2}}{\frac{t\Delta x}{2}}e^{-jt\left(x+\frac{\Delta x}{2}\right)}\varphi(t)dt. \tag{4.2.11}$$

由于

$$\left|\frac{\sin t\frac{\Delta x}{2}}{t\frac{\Delta x}{2}}e^{-jt\left(x+\frac{\Delta x}{2}\right)}\varphi(t)\right|\leqslant|\varphi(t)|$$

及 $\varphi(t)$绝对可积，故(4.2.11)式右端积分一致收敛．又由于在 t 的任一有限区间，(4.2.11)式右端被积函数一致收敛于 $e^{-jtx}\varphi(t)$．故

$$\begin{aligned}
&\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}\\
&=\lim_{\Delta x\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin t\frac{\Delta x}{2}}{t\frac{\Delta x}{2}}e^{-jt\left(x+\frac{\Delta x}{2}\right)}\varphi(t)dt
\end{aligned}$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{\Delta x\to 0}\frac{\sin t\frac{\Delta x}{2}}{t\frac{\Delta x}{2}}e^{-jt\left(x+\frac{\Delta x}{2}\right)}\varphi(t)dt$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-jtx}\varphi(t)dt. \qquad (4.2.12)$$

对于 $\Delta x<0$ 的情形，可类似证明得(4.2.12)式.

从(4.2.12)式看出 $\bar{F}(x)$导数存在，从而 $\bar{F}(x)$连续，故有 $\bar{F}(x)=F(x)$，并且

$$\bar{F}'(x)=F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-jtx}\varphi(t)dt.$$

用上述方法进一步可证明 $F'(x)$连续、有界且非负，因而有

$$F(x)=\int_{-\infty}^{x}F'(x)dx.$$

这样就得到了 ξ 为连续型且其密度函数为

$$f(x)=F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-jtx}\varphi(t)dt. \blacksquare$$

对于连续型随机变数 ξ，其特征函数 $\varphi(t)$是其密度函数 $f(x)$ 的富氏变换，而(4.2.10)式则表示 $f(x)$为 $\varphi(t)$的富氏反变换.

对于取整数的随机变数，有类似定理 4.2.1 的结果：

定理 4.2.2 设 ξ 为取整数值及 0 的随机变数，其概率函数为 $p_k=P\{\xi=k\}$, $k=\cdots-3,-2,-1,0,1,2,3,\cdots$，其特征函数为 $\varphi(t)=\sum_{k=-\infty}^{\infty}p_ke^{jtk}$，则

$$p_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-jtk}\varphi(t)dt. \qquad (4.2.13)$$

证 令 s 为任一整数，则有

$$\int_{-\pi}^{\pi}e^{-jts}\varphi(t)dt=\int_{-\pi}^{\pi}\sum_{k=-\infty}^{\infty}p_ke^{jtk}e^{-jts}dt$$

$$=\int_{-\pi}^{\pi}p_sdt+\sum_{\substack{k=-\infty\\k\neq s}}^{\infty}\int_{-\pi}^{\pi}p_ke^{jt(k-s)}dt.$$

因为当 $k \neq s$ 时，

$$\int_{-\pi}^{\pi} e^{jt(k-s)} dt = 0,$$

而

$$\int_{-\pi}^{\pi} p_s dt = 2\pi p_s$$

因而，

$$\int_{-\pi}^{\pi} e^{-jts} \varphi(t) dt = 2\pi p_s.$$

将 s 换为 k，并以 2π 除两边，得

$$p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-jtk} \varphi(t) dt. \quad ▌$$

我们将(4.1.2)式称为$\{p_k\}$的离散富氏变换，而上面的式子(4.2.13)则称为它的离散富氏反变换.

§4.3 相互独立随机变数和的特征函数

定理4.3.1 设$\xi_1, \xi_2, \cdots, \xi_n$为 n 个独立随机变数，令$\eta = \sum_{i=1}^{n} \xi_i$，则

$$\varphi_\eta(t) = \prod_{i=1}^{n} \varphi_{\xi_i}(t), \tag{4.3.1}$$

其中 $\varphi_\eta(t)$ 为 η 的特征函数，$\varphi_{\xi_i}(t)$ 为 ξ_i 的特征函数($i = 1, 2, \cdots, n$).

证 用归纳法证明.

令 $n = 2$，有

$$\begin{aligned} \varphi_\eta(t) &= E[e^{jt(\xi_1+\xi_2)}] = E[e^{jt\xi_1} \cdot e^{jt\xi_2}] \\ &= E(e^{jt\xi_1}) \cdot E(e^{jt\xi_2}) = \varphi_{\xi_1}(t) \cdot \varphi_{\xi_2}(t). \end{aligned} \tag{4.3.2}$$

设 $n = k$ 时，(4.3.1) 式成立，今证 $n = k+1$ 时，(4.3.1) 式也成立：

令 $\eta = \sum_{i=1}^{k+1} \xi_i = \sum_{i=1}^{k} \xi_i + \xi_{k+1}$，由(4.3.2)及假设，得

$$\varphi_\eta(t)=\varphi_{\sum\limits_{i=1}^{k}\xi_i}(t)\cdot\varphi_{\xi_{k+1}}(t)$$

$$=\prod_{i=1}^{k}\varphi_{\xi_i}(t)\cdot\varphi_{\xi_{k+1}}(t)=\prod_{i=1}^{k+1}\varphi_{\xi_i}(t).\blacksquare$$

下面介绍几个例子，说明定理 4.3.1 的应用.

例 4.3.1 设 $\xi_i(i=1, 2, \cdots, n)$ 为 n 个独立且同为服从两点分布的随机变数，根据定理 4.3.1, $\eta=\sum\limits_{i=1}^{n}\xi_i$ 的特征函数为

$$\varphi_\eta(t)=\prod_{i=1}^{n}\varphi_{\xi_i}(t)=(q+pe^{jt})^n.$$

上式恰好是二项分布的特征函数，由反演公式及唯一性定理知 η 服从二项分布.

应用例 4.3.1 我们进一步解释二项分布如何由贝努里概型产生.

令 ξ_i 表示贝努里试验的第 i 次试验的结果，事件 A 发生(成功)，记为 $\xi_i=1$，事件 A 不发生(失败)记为 $\xi_i=0$，则 $\eta=\sum\limits_{i=1}^{n}\xi_i$ 表示 n 次贝努里试验中事件 A 发生（成功）的总次数，它服从二项分布.

例 4.3.2 设 ξ_1, ξ_2 为相互独立且同为二项分布的随机变数，其概率函数为

$$P\{\xi_1=k\}=C_n^k p^k q^{n-k}\qquad k=1,2,\cdots,n,$$

$$P\{\xi_2=k\}=C_m^k p^k q^{m-k}\qquad k=1,2,\cdots,m.$$

由例 4.1.3 知

$$\varphi_{\xi_1}(t)=(q+pe^{jt})^n,$$

$$\varphi_{\xi_2}(t)=(q+pe^{jt})^m.$$

再由定理 4.3.1 得 $\eta=\xi_1+\xi_2$ 的特征函数

$$\varphi_\eta(t)=\varphi_{\xi_1}(t)\cdot\varphi_{\xi_2}(t)=(pe^{jt}+q)^{n+m}.$$

由分布函数与特征函数关系知上式右端是 $n+m$ 次贝努里试验的二项分布．这结果明显说明了如下两个事实：(1)两个独立的且具有二项分布的随机变数之和仍为二项分布（这结果对任意有穷个随机变数时也真）．(2)从直观上看，ξ_1 表示在 n 次贝努里试验中，某事件 A 出现次数，而 ξ_2 表示在另外 m 次贝努里试验中某事件 A 出现次数．若假定 $m+n$ 次试验是独立的，其和 $\eta=\xi_1+\xi_2$ 则表示 $m+n$ 次贝努里试验中某事件 A 出现的次数，故其分布也是二项分布．

例 4.3.3 设 ξ_1 和 ξ_2 分别是具有参数为 λ_1 和 λ_2 的泊松分布且相互独立的随机变数，由例 4.1.4 及定理 4.3.1 可得随机变数 $\eta=\xi_1+\xi_2$ 的特征函数为

$$\varphi_\eta(t)=[e^{\lambda_1(e^{jt}-1)}][e^{\lambda_2(e^{jt}-1)}]=e^{(\lambda_1+\lambda_2)(e^{jt}-1)},$$

即两个独立的具有泊松分布的随机变数之和，仍服从泊松分布，其参数为 $\lambda_1+\lambda_2$．这结果可推广到任意有穷个的情形．

例 4.3.4 设随机变数 ξ_1，ξ_2 相互独立且分别服从正态分布 $N(a_1,\sigma_1)$ 及 $N(a_2,\sigma_2)$ 则 $\eta=\xi_1+\xi_2$ 的特征函数为

$$\begin{aligned}\varphi_\eta(t)&=\varphi_{\xi_1}(t)\cdot\varphi_{\xi_2}(t)=e^{ja_1t+\frac{1}{2}\sigma_1^2t^2}\cdot e^{ja_2t+\frac{1}{2}\sigma_2^2t^2}\\&=e^{j(a_1+a_2)t+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2},\end{aligned}$$

即 η 为服从正态 $N(a_1+a_2,\sqrt{\sigma_1^2+\sigma_2^2})$ 的随机变数．

从例 4.3.2—例 4.3.4 中看到 n 个独立的具有二项分布或泊松分布或正态分布的随机变数之和的分布类型不变．这一性质，称之为"可加性"或"再生性"．这种性质对于服从 Γ-分布的随机变数也存在．

定理 4.3.1 的逆命题并不成立，即若 $\varphi_1(t),\varphi_2(t),\cdots,\varphi_n(t)$ 及 $\varphi(t)$ 分别为为随机变数 $\xi_1,\xi_2,\cdots,\xi_n$ 及 η 的特征函数，假设

$$\eta=\sum_{i=1}^{n}\xi_i,$$

$$\varphi(t)=\prod_{i=1}^{n}\varphi_i(t),$$

然而 $\xi_1,\xi_2,\cdots,\xi_n$ 不一定相互独立.

例 4.3.5 设随机变数 ξ 服从具有密度函数为 $\frac{1}{\pi}\cdot\frac{1}{1+x^2}$ 的哥西分布，由例 4.1.9 知其特征函数为 $\varphi_\xi(t)=e^{-|t|}$. 令 $\eta=\xi+\xi=2\xi$，则 η 的特征函数为

$$\varphi_\eta(t)=\varphi_{2\xi}(t)=e^{-2|t|}=\varphi_\xi(t)\cdot\varphi_\xi(t)$$

虽然 η 的特征函数 $\varphi_\eta(t)$ 等于 ξ 的特征函数的自乘，然而显然 ξ 与 ξ 并不是相互独立的.

§4.4 多维随机变数的特征函数

一维随机变数的特征函数的概念，可推广到有穷维随机变数的情形，下面只限于讨论二维时的情形.

一、定义及例

定义 4.4.1 设 (ξ_1,ξ_2) 是一个二维随机变数，其分布函数为 $F(x_1,x_2)$，t_1,t_2 为任意实数，记

$$\varphi(t_1,t_2)=E[e^{j(t_1\xi_1+t_2\xi_2)}]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{j(t_1x_1+t_2x_2)}dF(x_1,x_2),\tag{4.4.1}$$

称 $\varphi(t_1,t_2)$ 为 (ξ_1,ξ_2) 的特征函数.

与一维时一样，因为 $e^{j(t_1x_1+t_2x_2)}$ 有界，故二维随机变数 (ξ_1,ξ_2) 的特征函数总是存在的，并且当 (ξ_1,ξ_2) 为离散型随机变数时，

$$\varphi(t_1,t_2)=\sum_r\sum_s e^{j(t_1x_{1r}+t_2x_{2s})}p(r,s),\tag{4.4.2}$$

其中 $p(r,s)=P\{\xi_1=x_{1r},\xi_2=x_{2s}\}$.

当(ξ_1,ξ_2)为连续型随机变数时，

$$\varphi(t_1, t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{j(t_1x_1+t_2x_2)}f(x, x_2)\,dx\,dx_2. \qquad (4.4.3)$$

例 4.4.1 设二维随机变数的分布列为

$$P\{\xi_1=1,\ \xi_2=1\}=\frac{1}{3}, P\{\xi_1=1,\ \xi_2=-1\}=\frac{1}{3},$$

$$P\{\xi_1=-1,\ \xi_2=1\}=\frac{1}{6}, P\{\xi_1=-1,\ \xi_2=-1\}=\frac{1}{6}.$$

由(4.4.2)式计算得(ξ_1,ξ_2)的特征函数

$$\begin{aligned}\varphi(t_1, t_2)&=\frac{1}{3}e^{j(t_1+t_2)}+\frac{1}{3}e^{j(t_1-t_2)}+\frac{1}{6}e^{j(-t_1+t_2)}+\frac{1}{6}e^{j(-t_1-t_2)}\\&=\frac{1}{3}e^{jt_1}(e^{jt_2}+e^{-jt_2})+\frac{1}{6}e^{-jt_1}(e^{jt_2}+e^{-jt_2})\\&=\frac{1}{6}(e^{jt_2}+e^{-jt_2})(2e^{jt_1}+e^{-jt_1})\\&=\frac{1}{3}\cos t_2(3\cos t_1+j\sin t_1).\end{aligned}$$

例 4.4.2 设随机变数(ξ_1,ξ_2)服从二维正态$N(m_1,\ \sigma_1;\ m_2,\ \sigma_2;r)$，它的密度函数为

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}e^{-\frac{1}{2(1-r^2)}\left[\left(\frac{x_1-m_1}{\sigma_1}\right)^2-2r\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2}+\left(\frac{x_2-m_2}{\sigma_2}\right)^2\right]},$$

则它的特征函数为

$$(x_1,x_2)=$$

$$\begin{aligned}\varphi(t_1, t_2)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{j(t_1x_1+t_2x_2)}f(x_1,x_2)\,dx_1dx_2\\&=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(\frac{x_2-m_2}{\sigma_2}\right)^2}\cdot e^{j(t_1x_1+t_2x_2)}\\&\qquad\cdot e^{-\frac{1}{2(1-r^2)}\left[\frac{x_1-m_1}{\sigma_1}-r\frac{x_2-m_2}{\sigma_2}\right]^2}dx_1dx_2.\end{aligned}$$

作变换

$$u=\frac{1}{\sqrt{1-r^2}}\left(\frac{x_1-m_1}{\sigma_1}-r\frac{x_2-m_2}{\sigma_2}\right), v=\frac{x_2-m_2}{\sigma_2},$$

则

$$\varphi(t_1, t_2)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{j\{t_1[(u\sqrt{1-r^2}+rv)\sigma_1+m_1]+t_2(\sigma_2 v+m_2)\}}$$

$$\cdot e^{-\frac{u^2}{2}-\frac{v^2}{2}}dudv$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{j(t_1m_1+t_2m_2)}\cdot e^{jt_1\sigma_1\sqrt{1-r^2}u}e^{-\frac{1}{2}u^2}$$

$$\cdot e^{j(t_1\sigma_1 r+t_2\sigma_2)v}\cdot e^{-\frac{1}{2}v^2}dudv$$

$$=e^{j(t_1m_1+t_2m_2)}\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{jt_1(\sigma_1\sqrt{1-r^2})u}e^{-\frac{1}{2}u^2}du$$

$$\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{j(t_1\sigma_1 r+t_2\sigma_2)v}e^{-\frac{1}{2}v^2}dv$$

$$=e^{j(t_1m_1+t_2m)}e^{-\frac{1}{2}(1-r^2)\sigma_1^2t_1^2}\cdot e^{-\frac{1}{2}(t_1\sigma_1 r+t_2\sigma_2)^2}$$

$$=e^{j(t_1m_1+t_2m_2)}e^{-\frac{1}{2}(\sigma_1^2t_1^2+2\sigma_1\sigma_2 r t_1t_2+\sigma_2^2t_2^2)}.$$

特别，当 $m_1=m_2=0,\sigma_1=\sigma_2=1$ 时，则有

$$\varphi(t_1, t_2)=e^{-\frac{1}{2}(t_1^2+2rt_1t_2+t_2^2)}.$$

二、二维随机变数特征函数的性质

二维随机变数的特征函数，也有类似一维特征函数的性质，证明方法类似．下面只列性质，不一一详细证明．

性质 4.4.1 设随机变数 (ξ_1,ξ_2) 的特征函数为 $\varphi(t_1,t_2)$ 则

1) $\varphi(0,0)=1$，且对任意 $t_1,t_2\in R_1$

$$|\varphi(t_1,t_2)|\leqslant\varphi(0,0)=1;$$

2) $\varphi(-t_1, -t_2)=\overline{\varphi(t_1, t_2)}$;

3) $\varphi(t_1,t_2)$ 于实平面上一致连续；

4) $\varphi(t_1,0)=\varphi_1(t_1)$；$\varphi(0,t_2)=\varphi_2(t_2)$．

其中 $\varphi_1(t)$，$\varphi_2(t)$ 为 ξ_1 及 ξ_2 的特征函数．

例 4.4.3 如例 4.4.2，得

$$\varphi_1(t)=\varphi(t,0)=e^{jtm_1-\frac{1}{2}\sigma_1^2t^2},$$

$$\varphi_2(t)=\varphi(0,t)=e^{itm_2-\frac{1}{2}\sigma_2^2t^2}.$$

由唯一性定理知 $\varphi_1(t)$，$\varphi_2(t)$ 分别为正态 $N(m_1,\sigma_1)$ 及 $N(m_2,\sigma_2)$ 的特征函数. 这里我们又一次得到了在第二章所述的关于二维正态分布的边沿分布也是正态分布的结果.

性质 4.4.2 设 a_1,a_2,b_1,b_2 皆为常数，(ξ_1,ξ_2) 为二维随机变数，则随机变数 $(a_1\xi_1+b_1,a_2\xi_2+b_2)$ 的特征函数为

$$e^{i(t_1b_1+t_2b_2)}\varphi(a_1t_1,a_2t_2),$$

其中 $\varphi(t_1,t_2)$ 为 (ξ_1,ξ_2) 的特征函数.

例 4.4.4 设随机变数 (ξ_1,ξ_2) 如例 4.4.2，则根据性质 4.4.2，$\left(\frac{\xi_1-m_1}{\sigma_1},\frac{\xi_2-m_2}{\sigma_2}\right)$ 的特征函数为

$$e^{-i\left(\frac{t_1m_1}{\sigma_1}+\frac{t_2m_2}{\sigma_2}\right)}\cdot e^{i\left(\frac{t_1m_1}{\sigma_1}+\frac{t_2m_2}{\sigma_2}\right)-\frac{1}{2}\left[\sigma_1^2\left(\frac{t_1}{\sigma_1}\right)^2+2r\sigma_1\sigma_2\frac{t_1}{\sigma_1}\cdot\frac{t_2}{\sigma_2}+\sigma_2^2\left(\frac{t_2}{\sigma_2}\right)^2\right]}$$

$$=e^{-\frac{1}{2}(t_1^2+2rt_1t_2+t_2^2)}$$

性质 4.4.3 两个二元分布函数 $F_1(x_1,x_2)$，$F_2(x_1,x_2)$ 恒等的充分必要条件是它们的特征函数 $\varphi_1(t_1,t_2)$ 和 $\varphi_2(t_1,t_2)$ 恒等.

性质 4.4.4 随机变数 ξ_1 与 ξ_2 相互独立的充分必要条件为 (ξ_1,ξ_2) 的特征函数

$$\varphi(t_1,t_2)=\varphi_{\xi_1}(t_1)\cdot\varphi_{\xi_2}(t_2).$$

其中 $\varphi_{\xi_1}(t_1)$，$\varphi_{\xi_2}(t_2)$ 分别为 ξ_1 和 ξ_2 的特征函数.

性质 4.4.5 设随机变数 (ξ_1,ξ_2) 的特征函数为 $\varphi(t_1,t_2)$，a_1，a_2，b 为任意常数，则 $\eta=a_1\xi_1+a_2\xi_2+b$ 的特征函数为

$$\varphi_\eta(t)=e^{itb}\varphi(a_1t,a_2t),$$

特别

(1) ξ_1,ξ_2 相互独立时，有

$$\varphi_\eta(t)=e^{itb}\varphi_{\xi_1}(a_1t)\cdot\varphi_{\xi_2}(a_2t).$$

(2) 对于 $a_1=a_2=1$，$b=0$，则 $\eta=\xi_1+\xi_2$ 的特征函数

$$\varphi_\eta = \varphi(t, t).$$

例 4.4.5 仍以例 4.4.2 为例

$$\eta = \frac{\xi_1 - m_1}{\sigma_1} + \frac{\xi_2 - m_2}{\sigma_2} = \frac{1}{\sigma_1}\xi_1 + \frac{1}{\sigma_2}\xi_2 - \left(\frac{m_1}{\sigma_1} + \frac{m_2}{\sigma_2}\right)$$

的特征函数可由例 4.4.2(或例 4.4.4)及性质 4.4.5 求得为

$$\varphi_\eta(t) = e^{-\frac{1}{2}(2+2r)t^2} = e^{-(1+r)t^2}.$$

由反演公式及唯一性定理得知 η 服从正态分布，其数学期望为 0，方差为 $2(1+r)$.

例 4.4.5 说明，若 (ξ_1, ξ_2) 服从正态分布，则 ξ_1, ξ_2 的线性组合也服从正态分布，下面定理证明它的逆命题也是成立的.

定理 4.4.1 随机变数 (ξ_1, ξ_2) 服从二维正态分布的充分必要条件为 ξ_1 与 ξ_2 的任一线性组合

$$\lambda_1\xi_1 + \lambda_2\xi_2 + \lambda_0 \tag{4.4.4}$$

有一维正态分布. 其中 $\lambda_0, \lambda_1, \lambda_2$ 为任意实常数，且 λ_1, λ_2 不全为 0.

证 必要性可由例 4.4.5 相同方法证明. 现证明充分性. 在 (4.4.4) 式中，令 $\lambda_0 = \lambda_2 = 0, \lambda_1 = 1$，由假设知 $\eta = \xi_1$ 服从正态分布. 设 $E(\xi_1) = m_1, D(\xi_1) = \sigma_1^2$. 又令 $\lambda_0 = \lambda_1 = 0, \lambda_2 = 1$，由假设知 $\eta = \xi_2$ 服从正态分布. 又设 $E(\xi_2) = m_2, D(\xi_2) = \sigma_2^2$.

令 $\sigma_{12} = \sigma_{21} = E(\xi_1 - m_1)(\xi_2 - m_2)$，由上述知道 $\sigma_{12} = \sigma_{21}$ 是存在的，这样，按假设 $\eta = \lambda_1\xi_1 + \lambda_2\xi_2$ 服从正态分布且

$$E(\eta) = \lambda_1 m_1 + \lambda_2 m_2,$$

$$D(\eta) = \lambda_1^2\sigma_1^2 + \lambda_2^2\sigma_2^2 + 2\sigma_{12}\lambda_1\lambda_2$$

$$= \lambda_1^2\sigma_1^2 + \lambda_2^2\sigma_2^2 + 2r\sigma_1\sigma_2\lambda_1\lambda_2.$$

其中 r 为相关系数，故 η 的特征函数为

$$E(e^{jt\eta})=e^{jt(\lambda_1 m_1+\lambda_2 m_2)}\cdot e^{-\frac{1}{2}(\lambda_1^2\sigma_1^2+\lambda_2^2\sigma_2^2+2r\sigma_1\sigma_2\lambda_1\lambda_2)t^2}.$$

令 $t=1$ 得

$$E\left[e^{j(\lambda_1\xi_1+\lambda_2\xi_2)}\right]=e^{j(\lambda_1 m_1+\lambda_2 m_2)-\frac{1}{2}(\lambda_1^2\sigma_1^2+\lambda_2^2\sigma_2^2+2r\sigma_1\sigma_2\lambda_1\lambda_2)}. \tag{4.4.5}$$

由于 λ_1,λ_2 为任意实数，故可把 λ_1,λ_2 的位置当作二维随机变数的特征函数中的变元 t_1,t_2，由(4.4.5)式及根据性质 4.4.3 得知 (ξ_1,ξ_2) 服从二维正态 $N(m_1,m_2,\sigma_1,\sigma_2;\ r)$。这就得到了充分性的证明。▌

定理 4.4.2 设(ξ_1,ξ_2)为二维随机变数，$E(\xi_1^{k_1},\xi_2^{k_2})$存在，则其特征函数 $\varphi(t_1,t_2)$的偏导数$\dfrac{\partial^{(k_1+k_2)}\varphi(t_1,t_2)}{\partial t_1^{k_1}\partial t_2^{k_2}}$存在且

$$E(\xi_1^{k_1}\cdot\xi_2^{k_2})=j^{-(k_1+k_2)}\left[\frac{\partial^{(k_1+k_2)}\varphi(t_1,t_2)}{\partial t_1^{k_1}\partial t_2^{k_2}}\right]_{t_1=t_2=0}.$$

例 4.4.6 设(ξ_1,ξ_2)如例 4.4.2，依定理 4.4.2 有

$$\begin{aligned}E(\xi_1)&=j^{-1}\left[\frac{\partial\varphi(t_1,t_2)}{\partial t_1}\right]_{t_1=t_2=0}\\&=-j\left[(jm_1-(\sigma_1^2t_1+r\sigma_1\sigma_2t_2))\right.\\&\quad\left.\cdot e^{j(t_1m_1+t_2m_2)-\frac{1}{2}(\sigma_1^2t_1^2+2r\sigma_1\sigma_2t_1t_2+\sigma_2^2t_2^2)}\right]_{t_1=t_2=0}\\&=m_1.\end{aligned}$$

用同样方法计算可得

$$E(\xi_1^2)=j^{-2}\left[\frac{\partial^2\varphi(t_1,t_2)}{\partial t_1^2}\right]_{t_1=t_2=0}=\sigma_1^2+m_1^2,$$

$$E(\xi_2)=m_2;\quad E(\xi_2^2)=\sigma_2^2+m_2^2,$$

$$E(\xi_1\cdot\xi_2)=j^{-2}\left[\frac{\partial^2\varphi(t_1,t_2)}{\partial t_1\partial t_2}\right]_{t_1=t_2=0}=r\sigma_1\sigma_2+m_1m_2.$$

§4.5 母 函 数

在研究只取有穷或无穷非负整数值($k=0,1,2,\cdots$)的随机变数时，用母函数来代替特征函数较为方便.

定义 4.5.1 设随机变数 ξ 的分布列为

$$p_k=P\{\xi=k\},k=0,1,2,\cdots$$

记实变数 s 的实函数

$$\psi_\xi(s)=E(s^\xi)=\sum_k p_k s^k.\quad(-1\leqslant s\leqslant 1).\qquad(4.5.1)$$

称 $\psi_\xi(s)$ 为 ξ 的母函数，如不产生混乱，$\psi_\xi(s)$ 简记为 $\psi(s)$.

由于 $\psi(1)=\sum_k p_k=1$，因此(4.5.1)式右端的幂级数在 $|s|\leqslant 1$ 上绝对且一致收敛，所以任一取非负整数值的随机变数的母函数总是存在的，并且在$[-1,1]$上是一致连续的.

例 4.5.1 设 ξ 服从两点分布，依(4.5.1)式可求得其母函数

$$\psi(s)=q+ps.$$

例 4.5.2 设 ξ 服从二项分布，可算得其母函数为

$$\psi(s)=\sum_{k=0}^n C_n^k p^k(1-p)^{n-k}s^k=\sum_{k=0}^n C_n^k(ps)^k(1-p)^{n-k}$$
$$=(q+ps)^n.$$

例 4.5.3 设 ξ 服从泊松分布，则其母函数为

$$\psi(s)=\sum_{k=0}^\infty e^{-\lambda}\frac{\lambda^k}{k!}s^k=\sum_{k=0}^\infty e^{-\lambda}\frac{(\lambda s)^k}{k!}$$
$$=e^{-\lambda}e^{\lambda s}=e^{\lambda(s-1)}.$$

母函数也有类似特征函数的一些性质，仅列在下面，不作证明，读者可用类似于证明特征函数性质的方法证明之.

性质 4.5.1 $\psi(1)=1$ 且

$$|\psi(s)|\leqslant\psi(1)=1.$$

性质 4.5.2 (线性性质)

设 ξ 的母函数为 $\psi_\xi(s)$，则 $\eta=a\xi+b$ 的母函数为 $\psi_\eta(s)=s^b\psi_\xi(s^a)$，其中 a,b 为非负整数.

例 4.5.4 令 ξ 服从两点分布，则 $\eta=\xi+1$ 的母函数为

$$\psi_\eta(s)=s(q+ps)=qs+ps^2.$$

性质 4.5.3 有穷个独立随机变数和的母函数等于各个随机变数的母函数之乘积，即若 $\xi_1,\xi_2,\cdots,\xi_n$ 的母函数分别为 $\psi_1(s)$，$\psi_2(s),\cdots,\psi_n(s)$，则 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 的母函数为

$$\psi_\eta(s)=\prod_{k=1}^{n}\psi_k(s).$$

例 4.5.5 设 $\xi_1,\xi_2,\cdots,\xi_n$ 皆为二点分布，且相互独立，则 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 的母函数为

$$\psi_\eta(s)=\prod_{k=1}^{n}(q+ps)=(q+ps)^n.$$

性质 4.5.4 若随机变数的 n 阶矩存在，则其母函数 $\psi(s)$ 的 $k(k\leqslant n)$ 阶导数 $\psi^{(k)}(s)$ 存在 $(|s|\leqslant 1)$，且 ξ 的 $k(k\leqslant n)$ 阶矩可由母函数在 $s=1$ 的各阶导数表示，如

$$E(\xi)=\psi'(1);\quad E(\xi^2)=\psi''(1)+\psi'(1). \tag{4.5.2}$$

例 4.5.6 设 ξ 服从二项分布，由例 4.5.2 知

$$\psi(s)=(q+ps)^n.$$

由(4.5.2)式得

$$E(\xi)=[np(q+ps)^{n-1}]_{s=1}=np$$

$$\begin{aligned}E(\xi^2)&=\psi''(1)+\psi'(1)\\&=[n(n-1)p^2(q+ps)^{n-2}]_{s=1}+np\\&=np(np+q).\end{aligned}$$

由此可得到

$$D(\xi)=n^2p^2+npq-(np)^2=npq.$$

性质 4.5.5 (反演公式)

设随机变数 ξ 的分布列为 $p_k,(k=0,1,\cdots)$，母函数为

$$\psi(s)=\sum_k p_k s^k \quad (|s|\leqslant 1),$$

则分布列可由下式给出：

$$p_k=\frac{1}{k!}\psi^{(k)}(0).$$

证 由幂级数性质知(4.5.1)式右端可逐项求导数，故

$$\psi^{(k)}(s)=k!p_k+\sum_{m\geqslant k+1} m(m-1)(m-2)\cdots(m-k+1)p_m s^{m-k},$$

$\therefore \quad \psi^{(k)}(0)=k!p_k$ 由此得 $p_k=\frac{1}{k!}\psi^{(k)}(0)$. ∎

例 4.5.7 设 ξ 为泊松分布，由例 4.5.3 知其母函数为 $e^{\lambda(s-1)}$. 由性质 4.5.5 得

$$p_k=\frac{1}{k!}[\lambda^k e^{\lambda(s-1)}]_{s=0}=\frac{\lambda^k}{k!}e^{-\lambda}.$$

对于多维随机变数的母函数，可仿多维随机变数的特征函数一样讨论. 下面仅讨论二维时的情形.

定义 4.5.2 设(ξ_1,ξ_2)为二维取非负整数的随机变数，其分布列为

$$p(i,k)=P\{\xi_1=i,\xi_2=k\} \quad (i,k=0,1,2,\cdots).$$

记

$$\psi(s_1,s_2)=E(s_1^{\xi_1}\cdot s_2^{\xi_2})=\sum_i\sum_k p(i,k)s_1^i\cdot s_2^k,$$

其中$|s_1|\leqslant 1$, $|s_2|\leqslant 1$. 称 $\psi(s_1,s_2)$为(ξ_1,ξ_2)的母函数，有时称为双变数母函数. 对于双变数母函数，有如下性质：

性质 4.5.6 对于任意 $s_1\in[-1,1]$, $s_2\in[-1,1]$,

$$|\psi(s_1,s_2)|\leqslant\psi(1,1)=1.$$

性质 4.5.7 设 $\psi_1(s)$, $\psi_2(s)$分别为 ξ_1 及 ξ_2 的母函数，则有

特征函数和母函数的定义、性质一览表(例在第二章常用分布表中):

		特　征　函　数	母　　函　　数
定义	一维	$\varphi(t)=E(e^{jt\xi})=\int_{-\infty}^{\infty}e^{jtx}dF(x)$ $-\infty<t<\infty$	$\psi(s)=E(s^{\xi})=\sum_{k}p_k s^k,\ \|s\|\leqslant 1$
	二维	$\varphi(t_1,t_2)=E[e^{j(t_1\xi_1+t_2\xi_2)}]$ $=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{j(t_1x_1+t_2x_2)}dF(x_1,x_2)$ $-\infty<t_1,\ t_2<\infty$	$\psi(s_1,s_2)=E(s_1^{\xi_1}\cdot s_2^{\xi_2})$ $=\sum_{i}\sum_{k}p(i,k)s_1^i\cdot s_2^k$ $-1\leqslant s_1,\ s_2\leqslant 1$
性质	1)	$\|\varphi(t)\|\leqslant\varphi(0)=1$ $\varphi(-t)=\overline{\varphi(t)}$	$\|\psi(s)\|\leqslant\psi(1)=1$
	2)	$\varphi_{a\xi+b}(t)=e^{jbt}\varphi_{\xi}(at)$	$\psi_{a\xi+b}(s)=s^b\psi_{a\xi}(s)=s^b\psi_{\xi}(s^a)$
	3)	$\varphi(t)$于$(-\infty,\infty)$一致连续.	$\psi(s)$于$[-1,1]$一致连续.
	4)	若ξ的n阶矩存在,则 $E(\xi^k)=j^{-k}\varphi^{(k)}(0),k\leqslant n$	若ξ的各阶矩存在,则 $E(\xi)=\psi'(1),E(\xi^2)=\psi''(1)+\psi'(1),$ ………………
	5)	反演公式: $F(x)=\lim_{x_1\to-\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-jtx_1}-e^{-jtx}}{jt}\varphi(t)dt$	反演公式: $p_n=\frac{1}{n!}\psi^{(n)}(0)$
	6)	$\|\varphi(t_1,t_2)\|\leqslant\varphi(0,0)=1$	$\|\psi(s_1,s_2)\|\leqslant\psi(1,1)=1$
	7)	$\varphi(t,0)=\varphi_1(t)$ $\varphi(0,t)=\varphi_2(t)$ 其中 $\varphi_1(t)$, $\varphi_2(t)$分别是 ξ_1,ξ_2的特征函数	$\psi(s,1)=\psi_1(s)$ $\psi(1,s)=\psi_2(s)$ 其中 $\psi_1(s),\psi_2(s)$ 分别是 ξ_1,ξ_2 的母函数
	8)	若ξ_1与ξ_2相互独立,则对任意 $t_1,t_2\in(-\infty,\infty)$有 $\varphi(t_1,t_2)=\varphi_1(t_1)\cdot\varphi_2(t_2)$	若ξ_1与ξ_2相互独立,则对任意 $s_1,s_2\in[-1,1]$有 $\psi(s_1,s_2)=\psi_1(s_1)\psi_2(s_2)$
	9)	$\varphi(t,t)=\varphi_{\xi_1+\xi_2}(t)$, 且当$\xi_1$与$\xi_2$相互独立时有 $\varphi_{\xi_1+\xi_2}(t)=\varphi_1(t)\cdot\varphi_2(t)$	$\psi(s,s)=\psi_{\xi_1+\xi_2}(s)$, 且当$\xi_1$与$\xi_2$相互独立时有 $\psi_{\xi_1+\xi_2}(s)=\psi_{\xi_1}(s)\cdot\psi_{\xi_2}(s)$

$$\psi(s,1)=\psi_1(s);\ \psi_1(1,s)=\psi_2(s).$$

性质 4.5.8 若 ξ_1 与 ξ_2 独立，则对于一切 $s_1\in[-1,\ 1]$ 及 $s_2\in[-1,1]$，有

$$\psi(s_1,s_2)=\psi_1(s_1)\psi_2(s_2).$$

性质 4.5.9 随机变数 $\xi_1+\xi_2$ 的母函数为 $\psi(s,\ s)$特别当 ξ_1 与 ξ_2 独立时，$\xi_1+\xi_2$ 的母函数为 $\psi_1(s)\psi_2(s)$.

例 4.5.8 设$(\xi_1,\ \xi_2)$服从二维两点分布(参看例 2.2.2)，则其双变数母函数为

$$\begin{aligned}\psi(s_1,s_2)&=\sum_{i=0}^{1}\ \sum_{k=0}^{1}p(i,k)s_1^i s_2^k\\&=p(0,0)+p(0,1)s_2+p(1,0)s_1+p(1,1)s_1s_2\\&=q+ps_1s_2.\end{aligned}$$

由性质 4.5.7 得

$$\psi_1(s)=\psi(s,1)=q+ps,\ \psi_2(s)=\psi(1,s)=q+ps.$$

由性质 4.5.9 得知，若(ξ_1,ξ_2)服从两点分布且 ξ_1，ξ_2 相互独立，则 $\xi_1+\xi_2$ 的母函数为

$$\psi_1(s)\psi_2(s)=(q+ps)^2.$$

一般地可得到 n 为有穷多个时的结果为$(q+ps)^n$.

习　题

1. 若随机变数 ξ 服从几何分布：

$$P\{\xi=k\}=pq^k,\ k=0,1,\cdots,$$

其中 $0<p<1,q=1-p$. 求 ξ 的特征函数 $\varphi(t)$，$E(\xi)$ 及 $D(\xi)$.

2. 求证：若随机变数 ξ 的特征函数是

$$\frac{e^{jt}(1-e^{jnt})}{n(1-e^{jt})},$$

则 ξ 以概率 $\frac{1}{n}$ 取值 $1,2,\cdots,n$.

3. 求证：函数

$$\varphi_1(t)=\sum_{k=0}^{\infty}a_k\cos kt,$$

$$\varphi_2(t)=\sum_{k=0}^{\infty}a_k e^{ikt}$$

(其中 $a_k\geqslant 0$, $\sum_{k=0}^{\infty}a_k=1$) 是特征函数，并决定其相应的概率分布。

4. 求下列各随机变数的概率分布 其特征函数分别为：

a) $\cos t$; b) $\cos^2 t$.

5. 设随机变数 ξ 的分布函数为 $F(x)$, $F(x)$ 连续且严格单调，求：

a) $\eta=aF(\xi)+b$;

b) $\zeta=\ln F(\xi)$

的特征函数。

6. 设 $F(x)$ 是随机变数 ξ 的分布函数，$\varphi(t)$ 为 ξ 的特征函数。令

$$G(x)=\frac{1}{2h}\int_{x-h}^{x+h}F(y)\,dy,$$

其中 $h>0$ 为常数，则 $G(x)$ 的相应的特征函数为 $\frac{\sin ht}{ht}\varphi(t)$.

7. 求证：如果 $F(x)$ 是分布函数，而 $\varphi(t)$ 是它的相应的特征函数，则对任何 x，恒成立等式：

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\varphi(t)e^{-itx}dt=F(x+0)-F(x-0).$$

如果 x_ν 是 $F(x)$ 的跳跃点的横坐标，则

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\varphi(t)|^2dt=\sum_{\nu}\{F(x_\nu+0)-F(x_\nu-0)\}^2.$$

8. 证明：特征函数 $\varphi(t)$ 是实值的充要条件是其相应的分布函数 $F(x)$ 是对称的(即 $F(x)$ 满足：$F(x)=1-F(-x+0)$)。

9. 求证：对任何实特征函数 $\varphi(t)$，以下二不等式成立：

$$1-\varphi(2t)\leqslant 4(1-\varphi(t)),$$

$$1+\varphi(2t)\geqslant 2[\varphi(t)]^2.$$

10. 试证：满足下列各等式的连续函数 $\varphi(t)$ 是特征函数：

i) $\varphi(t)=\varphi(-t)$;

ii) $\varphi(t+2a)=\varphi(t)$;

iii) $\varphi(t)=\dfrac{a-t}{a} \quad (0\leqslant t\leqslant a)$,

其中 a 是正常数.

11. 证明：若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立，都服从正态 $N(0,1)$分布，则 $\dfrac{1}{\sqrt{n}}\sum\limits_{k=1}^{n}\xi_k$ 也服从正态 $N(0,1)$分布.

12. 已知二维正态随机变数(ξ_1,ξ_2)中 ξ_1,ξ_2 服从 $N(0,\sigma_i^2)$，$(i=1,2)$ $\mathrm{cov}(\xi_1,\xi_2)=\sigma_{12}$，用特征函数法求 $E[(\xi_1^2-\sigma_1^2)(\xi_2^2-\sigma_2^2)]$.

13. 已知随机变数(ξ,η)的密度函数

$$f(x,y)=Ce^{-[4(x-5)^2+2(x-5)(y-3)+5(y-3)^2]}$$

求常数 C，$E(\xi)$，$E(\eta)$及特征函数 $\varphi(t_1,t_2)$.

14. 设 ξ 的母函数为 $\psi(s)$. 求下列各数列所对应的母函数：

当$n=0,1,2,\cdots$时，

a) $q_n=P(\xi\leqslant n)$；

b) $q_n=P(\xi<n)$；

c) $q_n=P(\xi\geqslant n)$；

d) $q_n=P(\xi>n+1)$；

e) $q_n=P(\xi=2n)$.

15. 在贝努里试验序列中，设 u_n 是第 $n-1$ 次与第 n 次试验首先出现组合 SF（S 表示成功，F 表示失败）的概率. 求其母函数，数学期望及方差.

16. 对于贝努里试验序列，令 S_r 表示第 r 次失败之前成功出现的总次数. 试求其母函数及概率分布.

17. 继上题. 称直到第一次失败为一轮. 设 R 是直到第 ν 个成功的轮数. 证明：

$$P(R=r)=p^{\nu}q^{r-1}C_{\nu+r-2}^{r-1}.$$

并求 $E(R)$及 $D(R)$.

18. 设$\{\xi_k\}$是相互独立的随机变数序列，且每个 ξ_k 都以概率$\dfrac{1}{a}$取值 0，1，$\cdots$，$a-1$.

令

$$S_n=\sum_{k=1}^{n}\xi_k.$$

证明：S_n 的母函数是 $\psi(s)=\left\{\frac{1-s^a}{a(1-s)}\right\}^n$

19. 设 u_n 是 n 次贝努里试验的成功次数能被 3 除尽的概率. 求关于 u_n 的一个递推关系式，然后求 u_n 的母函数。

第五章　相互独立随机序列的极限定理

极限定理是概率论的基本理论之一，它在概率论与数理统计的理论研究和应用中都十分重要.

如前所述，概率论与数理统计是研究随机现象的统计规律性的科学，但随机现象的统计规律性只有在相同条件下进行大量重复试验或观察才呈现出来. 所谓一个事件发生的频率具有稳定性，是指当试验的次数无限增大时，在某种收敛意义下逼近某一定数，这就是所谓"大数定律". 正是由于有这一定律，所以在许多问题中"概率"这一概率论中最基本的概念就具有一定的客观意义. 同样，所谓某一试验可能发生的各种结果的频率分布情况近似某一分布(如测量误差的分布近似于正态分布)，也是从某种极限意义说的. 没有"中心极限定理"，我们无从解释上述这种现象. 联系着 n 次贝努里试验的二项分布，当试验次数 n 无限增大时近似于泊松分布这一类问题也是这样.

从下一章开始讲述的数理统计的理论和应用中，我们将看到有一个很重要、很基本的问题，就是我们总是要通过有限的子样概率性质来推断总体的概率性质，而有限子样的概率性质(如它的分布函数)有时是较难获得的，需要借助于它的极限性质来近似.

综上所述，有些人认为概率论的真正历史应从出现贝努里大数定律的时刻算起.

这一章里主要讲述相互独立随机变数序列最基本的两种类型的极限定理，即"大数定律"和"中心极限定理".

§5.1 大数定律

定义 5.1.1 设 $\xi_n(n=1,2,\cdots)$ 为概率空间 $(\Omega,\mathscr{F},P)$ 上定义的随机变数序列(简称随机序列)，若存在随机变数 ξ，使对任意 $\varepsilon>0$ 恒有

$$\lim_{n\to\infty}P\{|\xi_n-\xi|\geqslant\varepsilon\}=0, \tag{5.1.1}$$

或等价地

$$\lim_{n\to\infty}P\{|\xi_n-\xi|<\varepsilon\}=1, \tag{5.1.2}$$

则称随机序列 $\{\xi_n\}$ 依概率收敛于随机变数 ξ(ξ 也可以是一个常数)．并用下面的符号表示：

$$\lim_{n\to\infty}\xi_n=\xi \quad (P), \quad 或 \quad \xi_n\xrightarrow{P}\xi.$$

随机序列 $\{\xi_n\}$ 依概率收敛和数学分析中的序列收敛有很大的不同．首先，随机序列 $\{\xi_n\}$ 是 $(\Omega,\mathscr{F},P)$ 上的一个函数序列．当我们说随机序列 $\{\xi_n\}$ 依概率收敛于 ξ，是指对任意 $\varepsilon>0$，事件 $\{|\xi_n-\xi|\geqslant\varepsilon\}$ 发生的概率，当 n 无限增大时，它无限接近于 0．而当我们说序列 $\left\{\frac{1}{n}\right\}$ 趋于0，是指当 n 无限增大时，$\frac{1}{n}$ 无限接近于 0．其次，随机序列依概率收敛与函数序列收敛也不一样．

例 5.1.1 设 $\Omega=[0,1]$，依次把 $[0,1]$ 分为 $1,2,2^2,\cdots,2^n,\cdots$ 等分．令

$$A_0=[0,1];$$

$$A_1=\left[0,\frac{1}{2}\right),\ A_2=\left[\frac{1}{2},1\right];$$

$$A_3=\left[0,\frac{1}{4}\right),\ A_4=\left[\frac{1}{4},\frac{1}{2}\right),\ A_5=\left[\frac{1}{2},\frac{3}{4}\right),\ A_6=\left[\frac{3}{4},1\right];$$

……………………………………

并令

$$\xi_n(\omega)=\begin{cases}1 & 当\ \omega\in A_n,\\ 0 & 当\ \omega\bar{\in}A_n,\end{cases} \quad n=1,2,\cdots$$

又设

$$P(\omega\in A_n)=\text{区间} A \text{的长度}.$$

显然$\xi_n\xrightarrow{P}0$，这是因为，对任意$\varepsilon>0$，

$$P\{\omega:\ |\xi_n(\omega)|>\varepsilon\}=P\{\omega\in A_n\}\to 0$$

但对任一$\omega\in[0,1]$，对某些n值有$\omega\in A_n$，对另一些n值，$\omega\bar{\in}A_n$，故对ω而言，$\xi_n(\omega)$时而取0，时而取1的值，因而对任何$\omega\in[0,1]$，$\xi_n(\omega)$皆不收敛.

定义 5.1.2 设$\{\xi_n\}$为一随机序列，数学期望$E\{\xi_n\}$存在，令$\bar{\xi}_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i$，若

$$\lim_{n\to\infty}[\bar{\xi}_n-E(\bar{\xi}_n)]=0 \qquad (P), \tag{5.1.3}$$

则称随机序列$\{\xi_n\}$服从大数定律，或说大数法则成立.

下面我们首先证明这一章一开始说及的频率"稳定于"概率的贝努里大数定律.

定理 5.1.1 (贝努里大数定律)

设$\xi_n(n=1,2,\cdots)$为相互独立同分布的随机序列，且

$$P\{\xi_n=1\}=p,\quad P\{\xi_n=0\}=q,$$

其中$q=1-p$，$0<p<1$，则$\{\xi_n\}$服从大数定律，即若令$\bar{\xi}_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i$，则有

$$\lim_{n\to\infty}\bar{\xi}_n=p \qquad (P).$$

证 ∵ $$E(\bar{\xi}_n)=p,$$

$$D(\bar{\xi}_n)=\frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i)=\frac{1}{n^2}npq=\frac{1}{n}pq.$$

由车贝谢夫不等式(定理3.1.2)得，对任意ε有

$$0\leqslant P\{|\bar{\xi}_n-p|\geqslant\varepsilon\}\leqslant\frac{pq}{n\varepsilon^2},$$

故得

$$\lim_{n\to\infty}P\{|\bar{\xi}_n-p|\geqslant\varepsilon\}=0. \ ▌$$

回忆例4.3.1，$\bar{\xi}_n$ 表示在 n 次贝努里试验中事件 A 发生的频率 $\frac{\mu}{n}$. 这样，定理 5.1.1 给出了频率 $\frac{\mu}{n}$"稳定于"概率 p 的一个数学表达式，也就是说，它给出了频率趋于概率的含义.

我们常常还需要考虑在 n 次独立试验中，事件 A 发生的概率 p_k 随试验次数 k 的变化而变化的情形. 对这种试验概型，有如下定理.

定理 5.1.2 （泊松大数定律）

设 $\xi_n(n=1,2,\cdots)$ 为相互独立随机序列，

$$P\{\xi_n=1\}=p_n,\quad P\{\xi_n=0\}=q_n,$$

其中 $q_n=1-p_n$. 则 $\{\xi_n\}$ 服从大数定律.

证 由定理所设可得

$$E(\bar{\xi}_n)=\frac{1}{n}\sum_{i=1}^{n}p_i=\bar{p}_n,$$

$$D(\bar{\xi}_n)=\frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i)=\frac{1}{n^2}\sum_{i=1}^{n}p_iq_i\leqslant\frac{1}{4n}.$$

由车贝谢夫不等式得，对任意 $\varepsilon>0$ 有

$$0\leqslant P\{|\bar{\xi}_n-\bar{p}_n|\geqslant\varepsilon\}\leqslant\frac{D(\bar{\xi}_n)}{\varepsilon^2}\leqslant\frac{1}{4n\varepsilon^2},$$

故得

$$\lim_{n\to\infty}P\{|\bar{\xi}_n-\bar{p}_n|\geqslant\varepsilon\}=0.\ ∎$$

现在让我们来分析一下上述定理 5.1.1 及定理 5.1.2 的条件和结果. 由两个定理的假设条件都可得到

$$D(\bar{\xi}_n)\leqslant\frac{1}{4n}.$$

这样，当 n 充分大时，$D(\bar{\xi}_n)$ 可以任意地小. 由方差的直观意义表明，当 n 充分大时，$\bar{\xi}_n$ 集中取值于它自己的数学期望 $E(\bar{\xi}_n)$（在定理 5.1.1 里为 p，而在定理 5.1.2里为 $\bar{p}_n$），而 $|\bar{\xi}_n-E(\bar{\xi}_n)|\geqslant\varepsilon$

这一事件发生的可能性是很小的，这就是两个定理的结果．因此，只要 $\bar{\xi}_n$ 的方差当 n 充分大时能任意地小，则大数定律就可以成立．下面的定理说明了这一点．

定理 5.1.3 （车贝谢夫大数定律）

设 ξ_n，$n=1, 2, \cdots$，为相互独立随机序列，若 $E(\xi_n)=a_n$，$D(\xi_n)=\sigma_n^2\leqslant c<\infty$．则$\{\xi_n\}$服从大数定律．

读者可用上面两定理的证明方法自行证明．

前面三个定理都是假定 ξ_n 的二阶矩存在且方差一致有界的前提下成立大数定律．然而在许多问题中，特别在数理统计学中，往往不能满足上述要求，而仅知道随机序列$\{\xi_n\}$是相互独立同分布的．对于这种情形有下述定理．

定理 5.1.4 （辛钦大数定律）

设 $\xi_n, n=1,2,\cdots$，是相互独立同分布的随机序列，若 ξ_n有有穷的数学期望 a，则$\{\xi_n\}$服从大数定律．

***证** 对任意 $\delta>0$，定义两个随机序列如下：

$$\eta_i=\xi_i\cdot\chi_{\{|\xi_i|<n\delta\}}=\begin{cases}\xi_i & \text{当 } |\xi_i|<n\delta; \\ 0 & \text{当 } |\xi_i|\geqslant n\delta,\end{cases}$$

$$\zeta_i=\xi_i\cdot\chi_{\{|\xi_i|\geqslant n\delta\}}=\begin{cases}0 & \text{当 } |\xi_i|<n\delta; \\ \xi_i & \text{当 } |\xi_i|\geqslant n\delta.\end{cases}$$

$$(i=1,2,\cdots,n)$$

令

$$M=E|\xi_i|,$$
$$a_n=E(\eta_i)$$
$$(i=1,2,\cdots,n).$$

由定理条件可得：

1） $\xi_i=\eta_i+\zeta_i \qquad (i=1,2,\cdots,n)$；

2） $M<\infty, a_n<\infty \qquad (n=1,2\cdots)$．并且，当 $n\to\infty$ 时，

$$a_n = E(\eta_i) = \int_{-\infty}^{\infty} x \chi_{\{|x|<n\delta\}} dF_{\xi_i}(x)$$

$$= \int_{-n\delta}^{n\delta} x dF_{\xi_i}(x) \to a. \qquad (5.1.4)$$

3) 对 $i=1,2,\cdots,n$ 有

$$D(\eta_i) = E(\eta_i^2) - a_n^2 \leqslant E(\eta_i^2) = \int_{|x|<n\delta} |x|^2 dF_{\xi_i}(x)$$

$$\leqslant n\delta \int_{-\infty}^{\infty} |x| dF_{\xi_i}(x) \leqslant nM\delta. \qquad (5.1.5)$$

由假设及(5.1.5)式得

$$D\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(\eta_i) \leqslant \frac{1}{n^2} n^2 M\delta = M\delta. \qquad (5.1.6)$$

4) 由(5.1.4),(5.1.6)两式及车贝谢夫不等式得，当 n 充分大时

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i - a\right| \geqslant 2\varepsilon\right\}$$

$$= P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i - a_n + a_n - a\right| \geqslant 2\varepsilon\right\}$$

$$\leqslant P\left\{\left[\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i - a_n\right| + \left|a_n - a\right|\right] \geqslant 2\varepsilon\right\}$$

$$\leqslant P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i - a_n\right| \geqslant \varepsilon\right\}$$

$$\leqslant \frac{1}{\varepsilon^2} D\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)$$

$$\leqslant \frac{M\delta}{\varepsilon^2}. \qquad (5.1.7)$$

5) 对任意 $i=1,2,\cdots,n$，当 n 充分大时注意到 $E|\xi_i|<\infty$，因

而有不等式$\int_{|x|\geqslant n\delta}|x|dF_{\xi_i}(x)<\delta^2$对任意$\delta$成立，因此我们有

$$P\{\zeta_i\neq 0\}=P\{|\xi_i|\geqslant n\delta\}=\int_{|x|\geqslant n\delta}dF_{\xi_i}(x)$$

$$\leqslant\frac{1}{n\delta}\int_{|x|\geqslant n\delta}|x|dF_{\xi_i}(x)<\frac{1}{n}\delta.$$

故当n充分大时

$$P\left\{\sum_{i=1}^{n}\zeta_i\neq 0\right\}\leqslant\sum_{i=1}^{n}P\{\zeta_i\neq 0\}<\delta. \qquad (5.1.8)$$

最后，由(5.1.7)式及(5.1.8)式，我们得到当n充分大时，有

$$P\{|\xi_n-a|\geqslant 2e\}=P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i-a+\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right|\geqslant 2e\right\}$$

$$\leqslant P\left\{\left[\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i-a\right|+\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right|\right]\geqslant 2e\right\}$$

$$\leqslant P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\eta_i-a\right|\geqslant 2e\right\}+P\left\{\sum_{i=1}^{n}\zeta_i\neq 0\right\}$$

$$\leqslant\frac{M\delta}{e^2}+\delta.$$

由δ的任意性，因此，当n充分大时，对任意$e>0$，上式右端可任意地小．因而定理5.1.4得证．∎

例5.1.2 用概率方法证明维尔斯特拉斯定理，

设$f(x)$为闭区间$[a, b]$上任一连续函数，则存在多项式序列$N_n(x)$，$(n=1,2,\cdots\cdots)$于$[a,b]$一致收敛于$f(x)$.

解 不失一般性，我们在闭区间$[0,1]$上证明上述结论．令

$$N_n(x)=\sum_{k=0}^{n}C_n^k x^k(1-x)^{n-k}f\left(\frac{k}{n}\right).$$

显然有$N_n(0)=f(0)$，$N_n(1)=f(1)$．因此，当$x=0$或者$x=1$的

收敛问题已经解决．现只考虑 $x\in(0,1)$ 时的情形．

考虑贝努里概型如下：设在每次试验中事件 A 发生的概率均为某一固定值 $x(0<x<1)$．μ_n 为 n 次试验中事件 A 发生的次数，则有

$$P\{\mu_n=k\}=C_n^k x^k(1-x)^{n-k},\quad k=0,1,\cdots,n.$$

因此

$$E\left[f\left(\frac{\mu_n}{n}\right)\right]=\sum_{k=0}^{n} f\left(\frac{\mu_n}{n}\right)P\{\mu_n=k\}$$

$$=\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k} f\left(\frac{k}{n}\right)=N_n(x),$$

$$E\left[f(x)-f\left(\frac{\mu_n}{n}\right)\right]=\sum_{k=0}^{n} C_n^k x^k(1-x)^{n-k}\left[f(x)-f\left(\frac{k}{n}\right)\right]$$

$$=f(x)-N_n(x).$$

另一方面，由 §3.5 全数学期望公式得

$$E\left[f(x)-f\left(\frac{\mu_n}{n}\right)\right]$$

$$=P\left\{\left|\frac{\mu_n}{n}-x\right|<\delta\right\}\cdot E\left\{\left[f(x)-f\left(\frac{\mu_n}{n}\right)\right]\Big/\left[\left|\frac{\mu_n}{n}-x\right|<\delta\right]\right\}$$

$$+P\left\{\left|\frac{\mu_n}{n}-x\right|\geqslant\delta\right\}\cdot E\left\{\left[f(x)-f\left(\frac{\mu_n}{n}\right)\right]\Big/\left[\left|\frac{\mu_n}{n}-x\right|\geqslant\delta\right]\right\},$$

其中 $\delta>0$ 用如下方法选定：对任意 $\varepsilon>0$，由于 $f(x)$ 是连续函数，所以，存在 $\delta>0$，使当 $|x-y|<\delta$ 时

$$|f(x)-f(y)|<\frac{\varepsilon}{2},$$

其中 $x,y\in(0,1)$。

令

$$M=\sup_{0\leqslant x\leqslant 1}|f(x)|,$$

因此得到

$$|f(x)-N_n(x)|\leqslant\frac{\varepsilon}{2}+P\left\{\left|\frac{\mu_n}{n}-x\right|\geqslant\delta\right\}\cdot 2M.$$

由贝努里大数定律知上式右端第二项中,

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-x\right|\geqslant\delta\right\}=0$$

对 $x\in(0,1)$一致成立. 事实上,我们可以从贝努里大数定律的证明过程得到. 这样,我们证明了对一切 $x\in[0,1]$一致地有

$$\lim_{n\to\infty}N_n(x)=f(x).$$

*§5.2 强大数定律

在§5.1里,贝努里大数定律指出,频率$\frac{\mu}{n}$依概率收敛于 p,但那里还不能得到频率$\frac{\mu}{n}$无限接近概率 p 这一事实. 正如例5.1.1所指出的,即使$\{\xi_n\}$依概率趋于0,但还不能说 ξ_n 可以无限地接近于0. 因此,不论 n 多大,贝努里大数定律还不能排除$\left\{\frac{\mu}{n}\neq p\right\}$这一事件发生的可能性. 波雷尔首先证明了比贝努里大数定律更强的结果,这就是我们在这一节要讲的强大数定律.

在讲述强大数定律之前,先引入一些概念及引理.

定义5.2.1 设 $\xi_n(n=1,2,\cdots\cdots)$是一定义于相同的概率空间 $(\Omega,\mathscr{F},P)$ 上的随机序列. 若存在一随机变数 ξ(可以是一常数),使

$$P\{\lim_{n\to\infty}\xi_n=\xi\}=1,$$

我们称随机序列$\{\xi_n\}$以概率为1收敛于 ξ, 或说几乎处处收敛于 ξ. 并记为

$$\lim_{n\to\infty}\xi_n=\xi\qquad(P\text{—}a.s.),$$

或者简记为

$$\lim_{n\to\infty}\xi_n=\xi \quad (a.s.) \quad 或 \quad \xi_n\xrightarrow{a.s.}\xi.$$

定义 5.2.2 设$\{\xi_n\}$为一随机序列，并且有有穷的数学期望$E(\xi_n)$. 令$\bar{\xi}_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i$，若

$$\bar{\xi}_n-E(\bar{\xi}_n)\xrightarrow{a.s.}0.$$

则称$\{\xi_n\}$服从强大数定律.

引理 5.2.1 设$\{\xi_n\}$为一随机序列，ξ为一随机变数(可以是一常数). 令

$$A_{n,m}=\left\{|\xi_n-\xi|\geqslant\frac{1}{m}\right\},$$

则

$$\{\lim_{n\to\infty}\xi_n=\xi\}=\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}\bar{A}_{k,m}. \tag{5.2.1}$$

证 设$\omega\in\{\lim_{n\to\infty}\xi_n=\xi\}$，则对任一正整数$m$，存在正整数$N$，当$n\geqslant N$时，有

$$\omega\in\left\{|\xi_n-\xi|<\frac{1}{m}\right\},$$

故有

$$\omega\in\bigcap_{k\geqslant N}\left\{|\xi_k-\xi|<\frac{1}{m}\right\},$$

更有

$$\omega\in\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}\left\{|\xi_k-\xi|<\frac{1}{m}\right\},$$

由m的任意性得

$$\omega\in\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}\left\{|\xi_k-\xi|<\frac{1}{m}\right\},$$

即

$$\omega \in \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_{k,m},$$

这样，我们得到

$$\{\lim_{n\to\infty}\xi_n = \xi\} \subset \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_{k,m}.$$

用同样的方法可证：

$$\{\lim_{n\to\infty}\xi_n = \xi\} \supset \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_{k,m}.$$

综上所述，便得

$$\{\lim_{n\to\infty}\xi_n = \xi\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \bar{A}_{k,m}. \quad \blacksquare$$

引理 5.2.2 设$\{\xi_n\}$为一随机序列，ξ为一随机变数，下面的诸命题等价：

1) $\xi_n \xrightarrow{a.s.} \xi$;

2) $P\left\{\bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k\geq n}\left[|\xi_k - \xi| < \frac{1}{m}\right]\right\} = 1$，亦即

$$P\left\{\bigcup_{m=1} \bigcap_{n=1} \bigcup_{k\geq n}\left[|\xi_k - \xi| \geq \frac{1}{m}\right]\right\} = 0;$$

3) 对任意 $e > 0$，

$$P\left\{\bigcup_{n=1}^{\infty} \bigcap_{k\geq n}[|\xi_k - \xi| < e]\right\} = 1，亦即$$

$$P\left\{\bigcap_{n=1}^{\infty} \bigcup_{k\geq n}[|\xi_k - \xi| \geq e]\right\} = 0;$$

4) 对任意 $e > 0$，

$$\lim_{n\to\infty} P\left\{\bigcap_{k\geq n}[|\xi_k - \xi| < e]\right\} = 1，亦即$$

$$\lim_{n\to\infty} P\left\{\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant e]\right\}=0.$$

证 由于2)，3)，4)中的每两个式子所指的事件都是互逆事件，因此在下面的证明中，只需证明其中一个式子等价即可.

由引理 5.2.1 立即可得 1)与 2)等价. 下面证 2)与 3)等价.

任给 $e>0$，则当 $m>\frac{1}{e}$时，有

$$\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant e]\subset\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right]$$
$$\subset\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right],$$

故有

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant e]\right\}\leqslant P\left\{\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right]\right\}.$$

从而由 2)式的第二个式子得到 3)式的第二个式子成立. 反过来，由 3)的第二个式子可得，对任一正整数 m，有

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right]\right\}=0,$$

从而

$$P\left\{\bigcup_{m=1}^{\infty}\bigcap_{n=1}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right]\right\}$$
$$\leqslant\sum_{m=1}^{\infty}P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}\left[|\xi_k-\xi|\geqslant\frac{1}{m}\right]\right\}=0.$$

因此 2)式的第二式成立.

对于 3)与 4)的等价性，我们只需注意到如下事实即可：对任意 n，

$$\bigcap_{k\geqslant n}[|\xi_k-\xi|<e]\subset\bigcap_{k\geqslant n+1}[|\xi_k-\xi|<e],\text{或}$$

$$\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant\varepsilon]\supset\bigcup_{k\geqslant n+1}[|\xi_k-\xi|\geqslant\varepsilon],$$

并且应用概率的连续性定理，可得

$$\lim_{n\to\infty}P\left\{\bigcap_{k\geqslant n}[|\xi_k-\xi|<\varepsilon]\right\}=P\left\{\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}[|\xi_k-\xi|<\varepsilon]\right\},$$

或者

$$\lim_{n\to\infty}P\left\{\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant\varepsilon]\right\}=P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}[|\xi_k-\xi|\geqslant\varepsilon]\right\}.\blacksquare$$

引理 5.2.3　波雷尔-坎特里引理.

设 $A_1,A_2,\cdots$ 是概率空间 $(\Omega,\mathscr{F},P)$ 上的一事件序列，令

$$p_k=P\{A_k\}.$$

1) 若　$\sum_{k=1}^{\infty}p_k<\infty$

则

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}=0;$$

2) 若　$\sum_{k=1}^{\infty}p_k=\infty$，且各 A_k 相互独立，则

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}=1.$$

证

1) 当　$\sum_{k=1}^{\infty}p_k<\infty$ 时，则对任意 $\varepsilon>0$，有正整数 N 存在，使得

$$0\leqslant P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}\leqslant P\left\{\bigcup_{k\geqslant N}A_k\right\}\leqslant\sum_{k\geqslant N}P\{A_k\}$$

$$=\sum_{k\geqslant N}p_k<\varepsilon.$$

故 1)得证.

2) 因为　$\sum_{k=1}^{\infty}p_k=\infty$，故对任意 n，有 $\sum_{k=n}^{\infty}p_k=\infty$.　由 $A_k(k=$

$1,2,\cdots$)的相互独立性及利用不等式

$$1-x<e^{-x},$$

得到

$$0\leqslant P\left(\bigcap_{k\geqslant n}\bar{A}_k\right)=\prod_{k=n}^{\infty}P\{\bar{A}_k\}=\prod_{k=n}^{\infty}[1-P\{A_k\}]$$

$$\leqslant e^{-\sum\limits_{k=n}^{\infty}p_k}=0,$$

即对任意 $n=1,2,\cdots$均有

$$P\left\{\bigcap_{k\geqslant n}\bar{A}_k\right\}=0.$$

因而得

$$P\left\{\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}\bar{A}_k\right\}=0.$$

换句话说

$$P\left\{\bigcup_{n=1}^{\infty}\bigcap_{k\geqslant n}A_k\right\}=1. \blacksquare$$

(1) $P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}=0$ 等价于说"只有有限个 A_k 发生的概率为1".

(2) $P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}=1$ 等价于说"有无穷多个 A_k 发生的概率为1".

上述(1)和(2)读者可自行验证.

定理 5.2.1 波雷尔强大数定律.

设 $\xi_n(n=1,2,\cdots)$相互独立、同分布,且

$$P\{\xi_n=1\}=p,P\{\xi_n=0\}=q,$$

其中 $q=1-p$, $0<p<1$. 则$\{\xi_n\}$服从强大数定律.

证 首先我们证明如下一个事实:若 η 为任一随机变数,它的四阶矩存在,则对任意 $\varepsilon>0$ 有

$$P\{|\eta-E(\eta)|\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon^4}E[\eta-E(\eta)]^4. \qquad (5.2.2)$$

事实上，

$$E[\eta-E(\eta)]^4=\int_{-\infty}^{\infty}[y-E(\eta)]^4dF_\eta(y)$$

$$\geqslant\int_{|y-E(\eta)|\geqslant\varepsilon}[y-E(\eta)]^4dF_\eta(y)$$

$$\geqslant\varepsilon^4\int_{|y-E(\eta)|\geqslant\varepsilon}dF_\eta(y)$$

$$=\varepsilon^4P\{|\eta-E(\eta)|\geqslant\varepsilon\}.$$

现在回到定理 5.2.1 的证明. 对任一 $\varepsilon>0$, 令 $\eta=\bar{\xi}_n$, $A_n=\{|\bar{\xi}_n-p|\geqslant\varepsilon\}$，因此，我们得到 $E(\eta)=p$，且

$$P\{A_n\}\leqslant\varepsilon^{-4}E(\bar{\xi}_n-p)^4.$$

由第三章习题 26 可得

$$E(\bar{\xi}_n-p)^4=\frac{pq(p^3+q^3)}{n^3}+\frac{3(n-1)}{n^3}p^2q^2,$$

令 $p_n=P\{A_n\}$，因而

$$p_n=P\{A_n\}\leqslant\varepsilon^{-4}\left[\frac{pq(p^3+q^3)}{n^3}+\frac{3(n-1)}{n^3}p^2q^2\right].$$

可见，对任意 $\varepsilon>0$, $\sum_{n=1}^{\infty}p_n<\infty$. 由波雷尔-坎特里引理得，对任意 $\varepsilon>0$

$$P\left\{\bigcap_{n=1}^{\infty}\bigcup_{k\geqslant n}A_k\right\}=0,$$

由引理 5.2.2 得

$$\bar{\xi}_n\xrightarrow{a.s.}p. \ \blacksquare$$

在定理 5.2.1 中，若令 ξ_k 表示贝努里试验中与第 k 次试验相联系的随机变数，则定理 5.2.1告诉我们，$\lim_{n\to\infty}\frac{\mu}{n}=p$ 成立的概率

为1. 换句话说，发生$\left\{\lim\limits_{n\to\infty}\frac{\mu}{n}\neq p\right\}$这一事件的概率为0(当然我们还不能说$\frac{\mu}{n}$必然趋于$p$). 这样，我们进一步得到了频率"稳定于"概率这一论断的含义. 它比贝努里大数定律有更强的结果.

在§5.1中讨论大数定律时，我们得到了车贝谢夫大数定律，在那里若把大数定律一词换为强大数定律时也是正确的. 同时，还可得到更一般的结果. 这一结果由柯尔莫哥洛夫获得. 我们将在下面讨论. 为此，先介绍噶依克-瑞尼不等式.

引理 5.2.4 噶依克-瑞尼不等式

设ξ_n, $(n=1, 2, \cdots)$为一相互独立随机序列，对任意$n=1,2,\cdots$，有

$$E(\xi_n)=0, D(\xi_n)=\sigma_n^2<\infty. \tag{5.2.3}$$

又设$\{c_n:\ n=1,2,\cdots\}$为非增的正常数序列. 则对任意正整数m, n, $m<n$及$\varepsilon>0$

$$P\left\{\max_{m\leq k\leq n} c_k\left|\sum_{i=1}^{k}\xi_i\right|\geq\varepsilon\right\}\leq\frac{1}{\varepsilon^2}\left(c_m^2\sum_{i=1}^{m}\sigma_i^2+\sum_{i=m+1}^{n}c_i^2\sigma_i^2\right).$$

特别，当$m=1, c_k=1, k=1,2,\cdots$，并注意到$\left\{\left|\max\limits_{1\leq k\leq n}\sum\limits_{i=1}^{k}\xi_i\right|\geq\varepsilon\right\}\subset\left\{\max\limits_{1\leq k\leq n}\left|\sum\limits_{i=1}^{k}\xi_i\right|\geq\varepsilon\right\}$，立即得到§3.4的最后一段指出的柯尔莫哥洛夫不等式.

证 令

$$\left.\begin{aligned}\eta_k&=\sum_{i=1}^{k}\xi_i, \qquad (k=1,2,\cdots),\\ \eta&=\sum_{k=m}^{n-1}\eta_k^2(c_k^2-c_{k+1}^2)+c_n^2\eta_n^2,\end{aligned}\right\} \tag{5.2.4}$$

则由(5.2.3)得

$$E(\eta_k^2)=\sum_{i=1}^{k}\sigma_i^2.$$

因而

$$\begin{aligned}
E(\eta)&=\sum_{k=m}^{n-1}E(\eta_k^2)\cdot(c_k^2-c_{k+1}^2)+c_n^2E(\eta_n^2)\\
&=c_m^2E(\eta_m^2)+\sum_{k=m+1}^{n-1}c_k^2[E(\eta_k^2)-E(\eta_{k-1}^2)]\\
&\qquad-c_n^2E(\eta_{n-1}^2)+c_n^2E(\eta_n^2)\\
&=c_m^2\sum_{k=1}^{m}\sigma_k^2+\sum_{k=m+1}^{n-1}c_k^2\sigma_k^2+c_n^2\sigma_n^2\\
&=c_m^2\sum_{k=1}^{m}\sigma_k^2+\sum_{k=m+1}^{n}c_k^2\sigma_k^2. \qquad (5.2.5)
\end{aligned}$$

令 $A_m=\{\omega:c_m|\eta_m|\geqslant\varepsilon\}$, $A_k=\{\omega:c_i|\eta_i|<\varepsilon$,

$(m\leqslant i<k)$, $c_k|\eta_k|\geqslant\varepsilon\}$, $k=m+1,m+2,\cdots,n$

则 $A_k(k=m,m+1,\cdots,n)$ 互不相容，且事件

$$\{\omega:\max_{m\leqslant k\leqslant n}c_k|\eta_k|\geqslant\varepsilon\}=\bigcup_{k=m}^{n}A_k,$$

因此

$$P\{\omega:\max_{m\leqslant k\leqslant n}c_k|\eta_k|\geqslant\varepsilon\}=\sum_{k=m}^{n}P(A_k). \qquad (5.2.6)$$

定义随机变数如下：

$$\chi_k(\omega)=\begin{cases}1, & \text{若 } \omega\in A_k;\\ 0, & \text{若 } \omega\overline{\in}A_k\end{cases}$$

$(k=m,m+1,\cdots,n)$. 由于 A_k 互不相容，故

$$\sum_{k=m}^{n}\chi_k(\omega)\leqslant 1.$$

以 η 乘上式两端并求数学期望得

$$E(\eta)\geqslant\sum_{k=m}^{n}E[\eta\cdot\chi_k(\omega)].\tag{5.2.7}$$

对于 $k<i\leqslant n$,

$$\eta_i=\eta_k+\xi_{k+1}+\cdots+\xi_i.$$

上式两边平方后再乘以 $\chi_k(\omega)$ 并求数学期望得

$$E[\eta_i^2\cdot\chi_k(\omega)]=E[\eta_k^2\cdot\chi_k(\omega)]+E[(\xi_{k+1}+\cdots+\xi_i)^2\cdot\chi_k(\omega)]$$
$$+2E[\eta_k(\xi_{k+1}+\cdots+\xi_i)\chi_k(\omega)],\tag{5.2.8}$$

由于 $(\xi_{k+1}+\cdots+\xi_i)^2\cdot\chi_k(\omega)\geqslant0$,故

$$E[(\xi_{k+1}+\cdots+\xi_i)^2\cdot\chi_k(\omega)]\geqslant0,$$

又由于 $\eta_k\chi_k(\omega)$ 只与 $\xi_1,\xi_2,\cdots,\xi_k$ 有关,故 $\eta_k\cdot\chi_k(\omega)$ 与 $\xi_{k+1}+\cdots+\xi_i$ 独立,因此

$$E[\eta_k(\xi_{k+1}+\cdots+\xi_i)\chi_k(\omega)]$$
$$=E[\eta_k\chi_k(\omega)]\cdot\left[\sum_{r=k+1}^{i}E(\xi_r)\right]=0.$$

这样一来,由(5.2.8)式得,对 $k\leqslant i\leqslant n$ 有

$$E[\eta_i^2\cdot\chi_k(\omega)]\geqslant E[\eta_k^2\cdot\chi_k(\omega)].\tag{5.2.9}$$

注意到在 A_k 上 $|\eta_k|\geqslant\frac{\varepsilon}{c_k}$,由(5.2.9)式得,对 $k\leqslant i\leqslant n$ 有

$$E[\eta_i^2\cdot\chi_k(\omega)]\geqslant\frac{\varepsilon^2}{c_k^2}E[\chi_k(\omega)]=\frac{\varepsilon^2}{c_k^2}P(A_k).$$

由上式及(5.2.4)式得到,当 $m\leqslant k\leqslant n$ 时

$$E[\eta\cdot\chi_k(\omega)]=\sum_{r=m}^{n-1}E[\eta_r^2\cdot\chi_k(\omega)]\cdot(c_r^2-c_{r+1}^2)+c_n^2E[\eta_n^2\chi_k(\omega)]$$
$$\geqslant\sum_{r=k}^{n-1}E[\eta_r^2\chi_k(\omega)](c_r^2-c_{r+1}^2)+c_n^2E[\eta_n^2\chi_k(\omega)]$$
$$\geqslant\frac{\varepsilon^2}{c_k^2}P(A_k)\left[\sum_{r=k}^{n-1}(c_r^2-c_{r+1}^2)+c_n^2\right]$$

$$=\varepsilon^2 P(A_k). \tag{5.2.10}$$

由(5.2.7)式及(5.2.10)式得到

$$E(\eta)\geqslant\varepsilon^2\sum_{k=m}^{n}P(A_k). \tag{5.2.11}$$

由(5.2.5),(5.2.6)及(5.2.11)三式得

$$P\left\{\max_{m\leqslant k\leqslant n}c_k\left|\sum_{i=1}^{k}\xi_i\right|\geqslant\varepsilon\right\}=\sum_{k=m}^{n}P(A_k)\leqslant\frac{1}{\varepsilon^2}E(\eta)$$

$$\leqslant\frac{1}{\varepsilon^2}\left[c_m^2\sum_{k=1}^{m}\sigma_k^2+\sum_{k=m+1}^{n}c_k^2\sigma_k^2\right].\ ■$$

定理 5.2.2　柯尔莫哥洛夫判别法

设 $\xi_n(n=1,2,\cdots)$ 为一相互独立随机序列，若

$$\sum_{n=1}^{\infty}\frac{D(\xi_n)}{n^2}<\infty, \tag{5.2.12}$$

则 $\{\xi_n\}$ 服从强大数定律.

证　令

$$c_k=\frac{1}{k},$$

$$\eta_k=\sum_{i=1}^{k}[\xi_i-E(\xi_i)],$$

故

$$\frac{1}{k}\eta_k=\bar{\xi}_k-E(\bar{\xi}_k),\ \text{且}\ E(\eta_k)=0,\ D(\eta_k)=\sum_{i=1}^{k}D(\xi_i).$$

由引理 5.2.4 可得对任意正整数 $m,n,m<n$ 及 $\varepsilon>0$

$$P\{\max_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon\}=P\left\{\max_{m\leqslant k\leqslant n}\left|\frac{1}{k}\eta_k\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left[\frac{1}{m^2}\sum_{k=1}^{m}D(\xi_k)+\sum_{k=m+1}^{n}\frac{D(\xi_k)}{k^2}\right]. \tag{5.2.13}$$

令 $B_n=\{\max\limits_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon\}$，显然 $B_n\subset B_{n+1}\subset\cdots$. 由概率连续性定理及(5.2.13)式得

$$0\leqslant P\left\{\bigcup_{n\geqslant m}[\max_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon]\right\}$$

$$=\lim_{n\to\infty}P\{\max_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left[\frac{1}{m^2}\sum_{k=1}^{m}D(\xi_k)+\sum_{k=m+1}^{\infty}\frac{D(\xi_k)}{k^2}\right].\qquad(5.2.14)$$

因此，对任意 $\varepsilon>0$

$$\lim_{m\to\infty}P\left\{\bigcup_{n\geqslant m}[\max_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon]\right\}=0.\qquad(5.2.15)$$

(5.2.15)式等价于

$$|\bar{\xi}_n-E(\bar{\xi}_n)|\xrightarrow{a.s.}0,$$

事实上，因为

$$\bigcup_{n\geqslant m}[\max_{m\leqslant k\leqslant n}|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon]=\bigcup_{n\geqslant m}\left\{\bigcup_{k=m}^{n}[|\bar{\xi}_k-E(\bar{\xi}_k)|\geqslant\varepsilon]\right\}$$

$$=\bigcup_{n\geqslant m}[|\bar{\xi}_n-E(\bar{\xi}_n)|\geqslant\varepsilon].$$

并且由引理 5.2.2，则立即可得上述结果. ▌

(5.2.15)的获得是由于如下事实：

1) 由假设 $\sum\limits_{k=1}^{\infty}\frac{D(\xi_k)}{k^2}<\infty$，故当 $m\to\infty$ 时，

$$\sum_{k=m+1}^{\infty}\frac{D(\xi_k)}{k^2}\to0.$$

由此(5.2.14)式右边第二项趋于 0.

2) (5.2.14)式右边的第一项当 $m\to\infty$ 时趋于 0，这是由于如下数学分析中的一个结果：若级数 $\sum\limits_{k=1}^{\infty}\frac{a_k}{k}$ 收敛，则 $\frac{1}{n}\sum\limits_{k=1}^{n}a_k\to0$,（当 $n\to\infty$ 时）.

证 令 $S_n=\sum_{k=1}^{n}\frac{a_k}{k}$，$n=1,2,\cdots$. 而

$$\sum_{k=1}^{n}a_k=\sum_{k=1}^{n}k\frac{a_k}{k}=\sum_{k=1}^{n}k\left[\sum_{m=1}^{k}\frac{a_m}{m}-\sum_{m=1}^{k-1}\frac{a_m}{m}\right]$$

$$=\sum_{k=1}^{n}[kS_k-kS_{k-1}]=\sum_{k=1}^{n}kS_k-\sum_{k=1}^{n}kS_{k-1}$$

$$=nS_n-\sum_{k=1}^{n-1}S_k\text{①}.$$

故

$$\frac{1}{n}\sum_{k=1}^{n}a_k=S_n-\frac{1}{n}\sum_{k=1}^{n-1}S_k$$

$$=S_n-\frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{k=1}^{n-1}S_k\to 0\ (n\to\infty).\ \blacksquare$$

利用上述结果，若令 $a_k=\frac{D(\xi_k)}{k}$，则由 $\sum_{k=1}^{\infty}\frac{D(\xi_k)}{k^2}<\infty$ 可推得当 $m\to\infty$ 时，$\frac{1}{m}\sum_{k=1}^{m}\frac{D(\xi_k)}{k}\to 0$. 因而得到，当 $m\to\infty$ 时

$$\frac{1}{m^2}\sum_{k=1}^{m}D(\xi_k)\leqslant\frac{1}{m}\sum_{k=1}^{m}\frac{D(\xi_k)}{k}\to 0.$$

推论 设 $\xi_n(n=1,2,\cdots)$ 为相互独立随机序列，若存在常数 C，使得对任意 $n=1,2,\cdots$，

$$D(\xi_n)=\sigma_n<C,$$

则 $\{\xi_n\}$ 服从强大数定律.

证 由 $\sum_{n=1}^{\infty}\frac{D(\xi_n)}{n^2}<C\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty$ 及定理 5.2.2 立即得到. $\blacksquare$

---

① 我们规定 $S_0=0$，$\sum_{n=1}^{0}$ 意味着不作和.

在§5.1里，我们曾得到辛钦大数定律：它是考虑相互独立同分布时的大数定律，若将大数定律换为强大数定律也是成立的，这结果也由柯尔莫哥洛夫得到．下面先证一个引理．

引理 5.2.5 （截尾法）

设 $\xi_n(n=1,2,\cdots)$ 为具有有穷数学期望 a 且有相同分布函数 $F(x)$ 的随机序列．令

$$\xi_n^*=\xi_n\chi_{\{|\xi_n|<n\}}(\omega)\quad(n=1,2,\cdots),$$

若

$$\sum_{n=1}^{\infty}P\{\xi_n^*\neq\xi_n\}<\infty,\tag{5.2.16}$$

则 $\{\xi_n^*\}$ 服从强大数定律的充分必要条件为 $\{\xi_n\}$ 服从强大数定律．

证 令

$$\bar{\xi}_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\ \bar{\xi}_n^*=\frac{1}{n}\sum_{i=1}^{n}\xi_i^*,\ E(\xi_i)=a.$$

我们先证必要性：

$$\because\quad|\bar{\xi}_n-a|=|\bar{\xi}_n-\bar{\xi}_n^*+\bar{\xi}_n^*-E(\bar{\xi}_n^*)+E(\bar{\xi}_n^*)-a|$$

$$\leqslant|\bar{\xi}_n-\bar{\xi}_n^*|+|\bar{\xi}_n^*-E(\bar{\xi}_n^*)|+|E(\bar{\xi}_n^*)-a|.\tag{5.2.17}$$

(5.2.17)式的右端以概率为1地成立如下事实：第一项由(5.2.16)式并根据波雷尔-坎特里引理以及引理5.2.2得到当 $n\to\infty$ 时趋于0；而第二项由假设 $\{\xi_n^*\}$ 服从强大数定律得到当 $n\to\infty$ 时趋于0；第三项由下面的证明也趋于0．因为

$$\lim_{n\to\infty}E(\xi_n^*)\ \lim_{n\to\infty}\int_{|x|<n}x\,dF(x)=E(\xi_n)=a,$$

由熟知的关于收敛序列的性质（若 $\lim\limits_{n\to\infty}a_n=a$，则可推得 $\frac{1}{n}\sum\limits_{k=1}^{n}a_k\to a$）得，当 $n\to\infty$ 时

$$|E(\bar{\xi}_n^*)-a|=\left|\frac{1}{n}E\left(\sum_{k=1}^{n}\xi_k^*\right)-a\right|$$

$$=\left|\frac{1}{n}\sum_{k=1}^{n}E(\xi_k^*)-a\right|\longrightarrow 0.$$

这样一来，由(5.2.17)可得，当 $n\to\infty$ 时，以概率为 1 地成立

$$|\bar{\xi}_n-a|\to 0.$$

用同样的方法可证其充分性. ▌

定理 5.2.3 （柯尔莫哥洛夫定理）

设 $\xi_n(n=1,2,\cdots)$ 为相互独立且具有相同分布的随机序列，若 $E|\xi_n|<\infty$，则 $\{\xi_n\}$ 服从强大数定律.

证 令

$$\xi_n^*=\xi_n\cdot\chi_{\{|\xi_n|<n\}}(\omega).$$

由假设 $E|\xi_n|<\infty$，即 $\int_{-\infty}^{\infty}|x|dF(x)<\infty$，其中 $F(x)$ 为 ξ_n 的分布函数. 因此，按 R-S 积分定义得

$$\begin{aligned}D(\xi_n^*)&\leqslant E(\xi_n^*)^2\\&\leqslant\int_{-n}^{n}x^2dF(x)\\&\leqslant\sum_{k=0}^{n}(k+1)^2P\{k\leqslant|\xi_n|<k+1\}.\end{aligned}$$

因此，

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{D(\xi_n^*)}{n^2}&\leqslant\sum_{n=1}^{\infty}\sum_{k=0}^{n}\frac{(k+1)^2}{n^2}P\{k\leqslant|\xi_n|<k+1\}\\&=P\{0\leqslant|\xi_n|<1\}\sum_{n=1}^{\infty}\frac{1}{n^2}+\\&\quad+\sum_{k=1}^{\infty}(k+1)^2P\{k\leqslant|\xi_n|<k+1\}\sum_{n=k}^{\infty}\frac{1}{n^2}\end{aligned}$$

$$\leqslant 2P\{0\leqslant|\xi_n|<1\}+2\sum_{k=1}^{\infty}\frac{(k+1)^2}{k}P\{k\leqslant|\xi_n|<k+1\}①$$

$$\leqslant 2P\{0\leqslant|\xi_n|<1\}+2\sum_{k=1}^{\infty}\frac{(k+k)^2}{k}P\{k\leqslant|\xi_n|<k+1\}$$

$$=2P\{0\leqslant|\xi_n|<1\}+8\sum_{k=1}^{\infty}kP\{k\leqslant|\xi_n|<k+1\}$$

$$\leqslant 2P\{0\leqslant|\xi_n|<1\}+8\sum_{k=0}^{\infty}\int_{k\leqslant|x|<k+1}|x|\,dF(x)$$

$$\leqslant 2P\{0\leqslant|\xi_n|<1\}+8\int_{-\infty}^{\infty}|x|\,dF(x)<\infty.$$

由定理 5.2.2 得$\{\xi_n^*\}$服从强大数定律.

另一方面,

$$\sum_{n=1}^{\infty}P\{\xi_n^*\neq\xi_n\}=\sum_{n=1}^{\infty}P\{|\xi_n|\geqslant n\}$$

$$=\sum_{n=1}^{\infty}\sum_{k\geqslant n}P\{k\leqslant|\xi_n|<k+1\}$$

$$=\sum_{k=0}^{\infty}kP\{k\leqslant|\xi_n|<k+1\}<\infty.$$

由引理 5.2.5 得到$\{\xi_n\}$服从强大数定律. ▌

*§ 5.3 依概率收敛与以概率为 1 收敛的关系

在 § 5.1 和 § 5.2 中讨论的大数定律和强大数定律,得到了相似的一些定理,似乎这二个概念完全一样,其实不然.

例 5.3.1 设随机序列 $\{\xi_n\}$ 的分布列定义如下:对于 $n=$

---

① 利用$\sum_{n=k}^{\infty}\frac{1}{(n+1)^2}<\sum_{n=k}^{\infty}\int_n^{n+1}\frac{1}{x^2}dx=\int_k^{\infty}\frac{1}{x^2}dx=\frac{1}{k}$得

$$\sum_{n=k}^{\infty}\frac{1}{n^2}<\frac{1}{k^2}+\frac{1}{k}<\frac{2}{k}.$$

$1,2,\cdots$

$$P\left\{\xi_n=\frac{1}{n}\right\}=1-\frac{1}{n},P\{\xi_n=n+1\}=\frac{1}{n}.$$

并假定$\xi_n(n=1,2,\cdots)$相互独立.

我们可证$\{\xi_n\}$依概率收敛于0，但它不以概率为1收敛于0.事实上对任意$\varepsilon>0$,有

$$\lim_{n\to\infty}P\{|\xi_n|\geqslant\varepsilon\}=\lim_{n\to\infty}P\{\xi_n=n+1\}$$

$$=\lim_{n\to\infty}\frac{1}{n}=0,$$

即
$$\xi_n\xrightarrow{P}0.$$

但若令

$$A_n=\{\xi_n=n+1\},$$

则有

$$P\{A_n\}=\frac{1}{n}.$$

而

$$\sum_{n=1}^{\infty}P\{A_n\}=\sum_{n=1}^{\infty}\frac{1}{n}=\infty.$$

由波雷尔-坎特里引理和引理5.2.2可得$\{\xi_n\}$不以概率为1收敛于0.

例5.3.1说明，由依概率收敛不能推出以概率为1收敛，但逆命题是成立的.我们有如下定理.

定理5.3.1 若$\xi_n\xrightarrow{a.s.}\xi$,则必$\xi_n\xrightarrow{P}\xi$.

证 对任意n,因为有

$$\{|\xi_n-\xi|\geqslant\varepsilon\}\subset\bigcup_{k\geqslant n}\{|\xi_k-\xi|\geqslant\varepsilon\},$$

故

$$0 \leqslant P\{|\xi_n-\xi| \geqslant \varepsilon\} \leqslant P\left\{\bigcup_{k \geqslant n}[|\xi_k-\xi| \geqslant \varepsilon]\right\}. \tag{5.3.1}$$

根据引理 5.2.2，由假设$\{\xi_n\}$以概率为1收敛于ξ得到：(5.3.1)式右边当$n\to\infty$时趋于0. 这样一来，我们有

$$\lim_{n\to\infty} P\{|\xi_n-\xi| \geqslant \varepsilon\}=0,$$

即

$$\xi_n \xrightarrow{P} \xi. \blacksquare$$

§5.4 中心极限定理

我们在第二章引入正态分布时，曾经讲到有许多随机现象服从正态分布(如测量误差、射击偏差、…)是由于许多彼此没有什么相依关系、对随机现象谁也不能起突出影响而均匀地起到微小作用的随机因素共同作用(这些因素的叠加)的结果. 如果联系于这个随机现象的随机变数为ξ，则它可以看成为许多相互独立的起微小作用的因素ξ_k的总和$\sum_k \xi_k$，而这个总和服从或近似地服从正态分布. 对于这种现象我们还可再举一个有趣的例子加以说明.

例 5.4.1 (高尔顿钉板试验)

如图 5.4.1 中每一黑点表示钉在板上的一颗钉子，它们彼此的距离均相等，上一层的每一颗的水平位置恰好位于下一层的两颗正中间. 从入口处放进一个直径略小于两颗钉子之间的距离的小圆玻璃球，当小圆球向下降落过程中，碰到钉子后皆以$\frac{1}{2}$的概

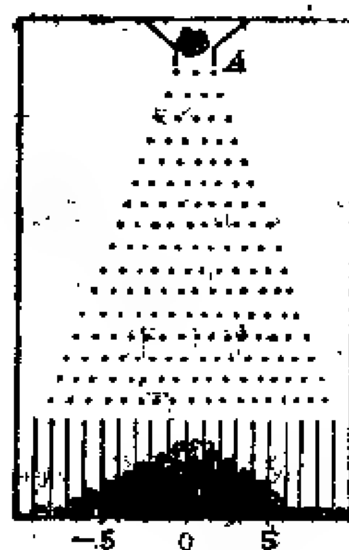


图 5.4.1

率向左或向右滚下，于是又碰到下一层钉子．如此继续下去，直到滚到底板的一个格子内为止．把许许多多同样大小的小球不断从入口处放下，只要球的数目相当大，它们在底板将堆成近似于正态$N(0,\sqrt{n})$的密度函数图形，示意图如图 5.4.1．其中 n 为钉子的层数．

我们作初步解释如下：

令 ξ_k 表示某一个小球在第 k 次碰了钉子后向左或向右落下这一随机现象相联系的随机变数（$\xi=1$ 表示向右落下，$\xi=-1$ 表示向左落下），由题意，ξ_k 的分布列可设为下述形式：对 $k=1,2,\cdots$

ξ_k	1	-1
p_k	$\frac{1}{2}$	$\frac{1}{2}$

则有，对 $k=1,2,\cdots$

$$E(\xi_k)=0, D(\xi_k)=1.$$

令

$$\zeta_n=\sum_{k=1}^{n}\xi_k,$$

其中 $\xi_k(k=1,2,\cdots,n)$ 相互独立．则 ζ_n 表示这个小球第 n 次碰钉后的位置．试验表明 $\zeta_n=\sum_{k=1}^{n}\xi_k$ 近似地服从正态分布．这一结果的数学证明将在以后给出．

上述例子说明，我们需要研究相互独立随机变数和的极限分布是正态分布的问题．这个问题的解决，对概率论在自然科学和技术应用中一个最重要的手段奠定了理论基础，这一手段是把一个现象或过程看作是许多因素的独立影响下出现的，而每一因素对该现象或过程所发生的影响都很小．如果我们关心的是该现象或过程的研究，而不是个别因素，则只要考虑这些因素的总的作

用就行了.

为研究上述提出的问题,先引入有关定义和定理.

一、依分布收敛

定义 5.4.1 设 $F_n(x)\ (n=1,\ 2,\ \cdots)$, $F(x)$ 分别为随机变数 $\xi_n\ (n=1,2,\cdots)$ 及 ξ 的分布函数, 若对于 $F(x)$ 的任一连续点 x 有

$$\lim_{n\to\infty}F_n(x)=F(x), \tag{5.4.1}$$

则称随机序列 $\{\xi_n\}$ 依分布收敛于 ξ, 并称 $F(x)$ 为 $\{F_n(x)\}$ 的极限分布函数.

如果对于分布函数列 $\{F_n(x)\}$ 存在一单调不减函数 $F(x)$, 使在 $F(x)$ 的每一连续点上 $\lim\limits_{n\to\infty}F_n(x)=F(x)$, 则称 $\{F_n(x)\}$ 弱收敛于 $F(x)$, 并记为:

$$\lim_{n\to\infty}F_n(x)=F(x) \quad (W) \quad 或 \quad F_n(x)\xrightarrow{W}F(x).$$

所以在定义 5.4.1 中我们是要求分布函数序列 $F_n\{x\}$ 弱收敛于分布函数 $F(x)$, 然而一般说来分布函数序列有可能弱收敛于一个不是分布函数的极限函数,这一点由下例即可看出.

例 5.4.2 考虑具有退化分布的随机序列 $\{\xi_n\}$, 它的分布列为

$$P\{\xi_n=n\}=1 \quad (n=1,2,\cdots).$$

这时

$$F_n(x)=\begin{cases}0, & 当\ x\leqslant n;\\ 1, & 当\ x>n,\end{cases} \quad (n=1,2,\cdots),$$

令

$$F(x)\equiv 0.$$

显然,对任意 $x\in R_1$ 有

$$\lim_{n\to\infty}F_n(x)=0=F(x),$$

因而有

$$F_n(x) \xrightarrow{W} F(x),$$

但 $F(x)$不是分布函数.

在定义5.4.1中,没有要求在$F(x)$的不连续点处的收敛性，否则,即使是处处收敛的随机序列也不能满足这一要求.

例5.4.3 设 $c_n(n=1, 2, \cdots)$为一严格单调上升收敛于 c 的常数序列,$c<\infty$.

在概率空间$(\Omega, \mathscr{F}, P)$上定义如下的随机变数:

令

$$\xi_n(\omega)=c_n \qquad (\text{对一切 } \omega\in\Omega),$$

$$\xi(\omega)=c \qquad (\text{对一切 } \omega\in\Omega),$$

且

$$P\{\xi_n(\omega)=c_n\}=1,$$

$$P\{\xi(\omega)=c\}=1.$$

显然,对每一 $\omega\in\Omega$ 有 $\lim\limits_{n\to\infty}\xi_n(\omega)=\xi(\omega)$,即$\{\xi_n(\omega)\}$处处收敛于 $\xi(\omega)$. 然而,ξ_n 的分布函数 $F_n(x)$ 不是点点收敛于 ξ 的分布函数 $F(x)$. 事实上，此时 $\xi_n(n=1, 2, \cdots)$及 ξ 的分布函数 $F_n(x)(n=1, 2, \cdots)$及 $F(x)$分别如下式所示:

$$F_n(x)=\begin{cases}0, \text{当 } x\leqslant c_n; \\ 1, \text{当 } x>c_n,\end{cases} \quad (n=1, 2, \cdots),$$

$$F(x)=\begin{cases}0, \text{当 } x\leqslant c; \\ 1, \text{当 } x>c.\end{cases}$$

显然,对 $F(x)$的连续点有$\lim\limits_{n\to\infty}F_n(x)=F(x)$, 但在 $F(x)$的不连续点 c,$F_n(c)$不收敛于 $F(c)$. 这是因为 $F_n(c)=1$,而 $F(c)=0$.

二、依分布收敛的充分必要条件

先考虑如下引理:

引理 5.4.1

任一分布函数序列是弱紧的. 即若 $\{F_n(x)\}$ 是一随机序列$\{\xi_n\}$的分布函数序列,则必存在$\{F_n(x)\}$的子序列$\{F_{n_k}(x)\}$弱收敛于有界不减、左连续函数 $F(x)$($F(x)$不一定是分布函数).

*证(对角线法) 令 $D=\{r_i\}$为全体有理数所组成的集合，因

为 $F_n(x)$是分布函数,故对 $n=1,2,\cdots$,有

$$0 \leqslant F_n(x) \leqslant 1.$$

由波尔察诺-维尔斯特拉斯定理知$\{F_n(r_1)\}$ 中必存在收敛子序列$\{F_{1n}(r_1)\}$. 同理,$\{F_{1n}(r_2)\}$中必存在收敛子序列$\{F_{2n}(r_2)\}\cdots$. 这样可得如下一系列的子序列:

$$\begin{array}{l} F_{11}(x),\quad F_{12}(x),\cdots \\ F_{21}(x),\quad F_{22}(x),\cdots \\ F_{31}(x),\quad F_{32}(x),\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}$$

在上述子序列中,每一子序列又为前一子序列的子序列. 考虑对角线上的函数 $F_{nn}(x)$作成的子序列$\{F_{nn}(x)\}$,显然,$\{F_{nn}(x)\}$对一切 $r_i \in D$ 都收敛($i=1,2,\cdots$).

设$\{F_{nn}(x)\}$的极限函数为 $F_D(x)$,即对任意 $x \in D$

$$\lim_{n \to \infty} F_{nn}(x) = F_D(x), \tag{5.4.2}$$

读者不难证明:$F_D(x)$是一个在D上有定义的有界 $(0 \leqslant F_D(x) \leqslant 1)$不减函数.

对任意 x,定义函数 $\bar{F}(x)$如下:

$$\bar{F}(x) = \sup_{\substack{r_i \leqslant x \\ r_i \in D}} F_D(r_i). \tag{5.4.3}$$

则我们得到了一个由 $F_D(x)$扩张成的,对任意 $x \in R_1$ 有定义的函数 $\bar{F}(x)$. 此函数 $\bar{F}(x)$仍为有界$(0 \leqslant \bar{F}(x) \leqslant 1)$、不减函数,且当 $x \in D$ 时,$\bar{F}(x) = F_D(x)$. $\bar{F}(x)$的上述性质,读者可以自行证明. 下面我们证明:

$$F_{nn}(x) \xrightarrow{W} \bar{F}(x).$$

设 x 是 $\bar{F}(x)$的连续点,由实数理论知道,存在二序列$\{r_i'\}$及

$\{r_i''\}$，$r_i', r_i'' \in D$，使得 $r_i' < x < r_i''$，且当 $r_i' \nearrow x$，$r_i'' \searrow x$ 时[①]，

$$\bar{F}(r_i'') - \bar{F}(r_i') \longrightarrow 0 \tag{5.4.4}$$

由于 $F_{nn}(x)$ 是分布函数，故对任意 $i=1,2,\cdots$

$$F_{nn}(r_i') \leqslant F_{nn}(x) \leqslant F_{nn}(r_i'').$$

由(5.4.2)式及上式，得

$$\begin{aligned}\bar{F}(r_i') &= F_D(r_i') = \lim_{n\to\infty} F_{nn}(r_i') \\ &= \lim_{n\to\infty}\inf F_{nn}(r_i') \leqslant \lim_{n\to\infty}\inf F_{nn}(x) \\ &\leqslant \lim_{n\to\infty}\sup F_{nn}(x) \leqslant \lim_{n\to\infty}\sup F_{nn}(r_i'') \\ &= F_D(r_i'') = \bar{F}(r_i'').\end{aligned}$$

由(5.4.4)式及 x 是 $\bar{F}(x)$ 的连续点得

$$\lim_{n\to\infty}\inf F_{nn}(x) = \lim_{n\to\infty}\sup F_{nn}(x) = \lim_{r_i'\nearrow x}\bar{F}(r_i') = \bar{F}(x),$$

即对任一 $\bar{F}(x)$ 的连续点 x，有

$$\lim_{n\to\infty} F_{nn}(x) = \bar{F}(x).$$

亦即我们证明了

$$F_{nn}(x) \xrightarrow{W} \bar{F}(x).$$

上述的极限函数 $\bar{F}(x)$ 还不一定是左连续的，但我们可以修改 $\bar{F}(x)$ 在不连续点的值，使之成为左连续的．我们可以如下进行修改：

令 A 表示 $\bar{F}(x)$ 的所有不连续点 x 所成的集合．定义 $F(x)$ 如下：

$$F(x) = \begin{cases} \bar{F}(x) & \text{当 } x \overline{\in} A \text{ 时;} \\ \lim\limits_{\substack{x_n \nearrow x \\ x_n \overline{\in} A}} \bar{F}(x_n) & \text{当 } x \in A \text{ 时.} \end{cases}$$

经修改后的函数 $F(x)$ 仍为有界（$0 \leqslant F(x) \leqslant 1$），不减，且左连续，$F_{nn}(x) \xrightarrow{W} F(x)$.

---

① 记号↗表示"单调上升趋于"，记号↘表示"单调下降趋于"，以下均同．

这样，$F_{nn}(x)(n=1,2,\cdots)$即为我们所求. ∎

引理 5.4.2 设 $F_n(x)(n=1,2,\cdots)$为分布函数序列，在有穷区间$[a,b]$上弱收敛于函数 $F(x)$，这里 a 和 b 都是 $F(x)$的连续点. 则对任一 $t\in R_1$，有

$$\lim_{n\to\infty}\int_a^b e^{jtx}dF_n(x)=\int_a^b e^{jtx}dF(x),$$

且对 t 的任一有穷区间$[T_1,T_2]$是一致的.

*证 对任意 $\varepsilon>0$，在$[a,b]$中总可选取分点 x_k:

$$a=x_0<x_1<\cdots<x_M=b,$$

使得每一 x_k 都是 $F(x)$的连续点，并且，

$$l=\max_{1\leqslant k\leqslant M}(x_k-x_{k-1})<\frac{\varepsilon}{3T},$$

其中 $T=\max\{|T_1|,\ |T_2|\}$. 令 $A_k=[x_{k-1},\ x_k)$ $(k=1,\ 2,\ \cdots,\ M-1)$，$A_M=[x_{M-1},x_M]$则

$$\begin{aligned}&\left|\int_a^b e^{jtx}dF_n(x)-\int_a^b e^{jtx}dF(x)\right|\\
&\leqslant\left|\sum_{k=1}^{M}\int_{A_k}e^{jtx_k}dF_n(x)-\sum_{k=1}^{M}\int_{A_k}e^{jtx_k}dF(x)\right|\\
&+\left|\sum_{k=1}^{M}\left[\int_{A_k}(e^{jtx_k}-e^{jtx})dF_n(x)-\int_{A_k}(e^{jtx_k}-e^{jtx})dF(x)\right]\right|\\
&\leqslant\sum_{k=1}^{M}\left|\int_{A_k}e^{jtx_k}dF_n(x)-\int_{A_k}e^{jtx_k}dF(x)\right|\\
&+\sum_{k=1}^{M}\left|\int_{A_k}(e^{jtx_k}-e^{jtx})dF_n(x)\right|\\
&+\sum_{k=1}^{M}\left|\int_{A_k}(e^{jtx_k}-e^{jtx})dF(x)\right|.\end{aligned}\tag{5.4.5}$$

由于

$$|e^{jtx_k}-e^{jtx}|=|e^{jt(x_k-x)}-1|=\left|jt\int_0^{x_k-x}e^{jty}dy\right|$$

$$\leqslant |t(x_k - x)|,$$

因此，对 $t \in [T_1, T_2]$ 有

$$\sum_{k=1}^{M} \left| \int_{\Delta_k} (e^{jtx_k} - e^{jtx}) dF_n(x) \right|$$

$$\leqslant \sum_{k=1}^{M} lT \int_{\Delta_k} dF_n(x) < \frac{\varepsilon}{3}. \qquad (5.4.6)$$

同理

$$\sum_{k=1}^{M} \left| \int_{\Delta_k} (e^{jtx_k} - e^{jtx}) dF(x) \right| < \frac{\varepsilon}{3}. \qquad (5.4.7)$$

另一方面

$$\left| \int_{\Delta_k} e^{jtx_k} dF_n(x) - \int_{\Delta_k} e^{jtx_k} dF(x) \right|$$

$$= \left| e^{jtx_k} [F_n(x_k) - F_n(x_{k-1}) - F(x_k) + F(x_{k-1})] \right|$$

$$\leqslant |F_n(x_k) - F(x_k)| + |F_n(x_{k-1}) - F(x_{k-1})|, \quad (5.4.8)$$

由于每一个 x_k 都是 $F(x)$ 的连续点，由假设，必存在正整数 N，当 $n \geqslant N$ 时，对 $k=1, 2, \cdots M$，皆有

$$|F_n(x_k) - F(x_k)| < \frac{\varepsilon}{6M}. \qquad (5.4.9)$$

将(5.4.9)代入(5.4.8)得

$$\left| \int_{\Delta_k} e^{jtx_k} dF_n(x) - \int_{\Delta_k} e^{jtx_k} dF(x) \right| < \frac{\varepsilon}{3M}. \qquad (5.4.10)$$

最后将(5.4.6)，(5.4.7)及(5.4.10)三式代入(5.4.5)式得到，对 $t \in [T_1, T_2]$ 一致地有

$$\left| \int_a^b e^{jtx} dF_n(x) - \int_a^b e^{jtx} dF(x) \right| < \varepsilon. \qquad (5.4.11)$$

由于 $[T_1, T_2]$ 的任意性，故知，对 $t \in R_1$，(5.4.11) 式亦成立. 因此引理 5.4.2 得证. ▮

有了上述两个引理，我们可以得到在以后研究中心极限定理时很有用的勒维-克拉美定理.

定理 5.4.1 勒维-克拉美定理

设 $F_n(x)$，$\varphi_n(t)$ 分别为随机序列 ξ_n 的分布函数和特征函数$(n=1,2,\cdots)$. 则 $\{\xi_n\}$ 依分布收敛于 ξ 的充分必要条件为，对每一 $t\in R_1$，$\lim\limits_{n\to\infty}\varphi_n(t)=\varphi(t)$，且对任一有限区间$[T_1,T_2]$是一致的. 其中 $\varphi(t)$ 为 ξ 的特征函数.

证 必要性证明如下:

令$\varphi_n(t)=\int_{-\infty}^{\infty}e^{jtx}dF_n(x)$, $\varphi(t)=\int_{-\infty}^{\infty}e^{jtx}dF(x)$. 其中 $F(x)$ 为 ξ 的分布函数.

设x',x''为 $F(x)$的连续点,不妨设 $x'<0$, $x''>0$. 则

$$
\begin{aligned}
&|\varphi_n(t)-\varphi(t)|\\
&=\left|\int_{-\infty}^{\infty}e^{jtx}dF_n(x)-\int_{-\infty}^{\infty}e^{jtx}dF(x)\right|\\
&\leqslant I_1+I_2+I_3,
\end{aligned}
\tag{5.4.12}
$$

其中

$$
\begin{aligned}
I_1&=\left|\int_{-\infty}^{x'}e^{jtx}dF_n(x)-\int_{-\infty}^{x'}e^{jtx}dF(x)\right|,\\
I_2&=\left|\int_{x'}^{x''}e^{jtx}dF_n(x)-\int_{x'}^{x''}e^{jtx}dF(x)\right|,\\
I_3&=\left|\int_{x''}^{\infty}e^{jtx}dF_n(x)-\int_{x''}^{\infty}e^{jtx}dF(x)\right|.
\end{aligned}
$$

下面逐一考虑I_1,I_2,I_3:

$$
\begin{aligned}
I_1&\leqslant\left|\int_{-\infty}^{x'}e^{jtx}dF_n(x)\right|+\left|\int_{-\infty}^{x'}e^{jtx}dF(x)\right|\\
&\leqslant\int_{-\infty}^{x'}dF_n(x)+\int_{-\infty}^{x'}dF(x)\\
&=F_n(x')+F(x').
\end{aligned}
$$

由分布函数性质及题设知，对任意 $\varepsilon>0$ 存在 x' 及正整数 N_1，当 $n\geqslant N_1$ 时，有

$$F_n(x')<\frac{\varepsilon}{6},\qquad F(x')<\frac{\varepsilon}{6}.$$

因此，对于 x'，当 $n\geqslant N_1$ 时

$$I_1<\frac{\varepsilon}{3}. \tag{5.4.13}$$

同理，存在正整数 N_2，使得对于 x''，当 $n\geqslant N_2$ 时

$$I_3<\frac{\varepsilon}{3}. \tag{5.4.14}$$

对于满足前述条件的 x' 和 x''，由引理5.4.2得，当 $n\geqslant N_3$ 时

$$I_2<\frac{\varepsilon}{3}. \tag{5.4.15}$$

最后，由(5.4.12)，(5.4.13)，(5.4.14)及(5.4.15)四式得到，当 $n\geqslant\max\{N_1,N_2,N_3\}$ 时，对 $t\in R_1$ 有

$$|\varphi_n(t)-\varphi(t)|<\varepsilon,$$

并且注意到引理 5.4.2 的结果，上述结果对 t 的任一有限区间 $[T_1,T_2]$ 是一致的.

下面证明充分性.

由引理5.4.1知道，$\{F_n(x)\}$ 有子序列 $\{F_{n_k}(x)\}$ 弱收敛于一个有界 $(0\leqslant F(x)\leqslant 1)$、不减、左连续函数 $F(x)$. 可以证明 $F(x)$ 为一分布函数，并且

$$\varphi(t)=\int_{-\infty}^{\infty}e^{itx}dF(x).$$

后者由必要性的证明直接可得. 因此，只要我们证明了 $F(x)$ 是分布函数即可. 即只须证

$$F(\infty)=1,F(-\infty)=0.$$

我们用反证法证明上述结果. 为此，令

$$\alpha=F(\infty)-F(-\infty)<1. \tag{5.4.16}$$

由假设$\varphi(t)$是一特征函数，故$\varphi(t)$是t的连续函数，并且$\varphi(0)=1$. 因此，存在 0 点的一个充分小的邻域$(-\delta,\delta)$，使得对满足不等式：$0<\varepsilon<1-\alpha$之ε，有

$$\frac{1}{2\delta}\left|\int_{-\delta}^{\delta}\varphi(t)\,dt\right|>1-\frac{\varepsilon}{2}>\alpha+\frac{\varepsilon}{2}. \qquad (5.4.17)$$

另一方面，既然$F_{n_k}(x)\xrightarrow[c]{W}F(x)$，以及由(5.4.16)，可取$x'>\frac{4}{\varepsilon\delta}$，使得$-x'$及$x'$皆为$F(x)$的连续点，并取正整数$k_0$，使得对于一切$k\geqslant k_0$，有

$$\alpha_k=F_{n_k}(x')-F_{n_k}(-x')<\alpha+\frac{\varepsilon}{4}.$$

由于$\varphi_{n_k}(t)$是$F_{n_k}(x)$的特征函数，故对上述δ及x'，有

$$\begin{aligned}\left|\int_{-\delta}^{\delta}\varphi_{n_k}(t)dt\right|&\leqslant\int_{-\infty}^{\infty}\left|\int_{-\delta}^{\delta}e^{jtx}dt\right|dF_{n_k}(x)\\&\leqslant\int_{|x|\leqslant x'}\left|\int_{-\delta}^{\delta}e^{jtx}dt\right|dF_{n_k}(x)\\&\quad+\int_{|x|>x'}\left|\int_{-\delta}^{\delta}e^{jtx}dt\right|dF_{n_k}(x).\end{aligned}$$

因为$|e^{jtx}|=1$，故

$$\left|\int_{-\delta}^{\delta}e^{jtx}dt\right|\leqslant 2\delta,$$

而当$|x|>x'$时，

$$\begin{aligned}\left|\int_{-\delta}^{\delta}e^{jtx}dt\right|&=\left|\frac{2}{x}\sin\delta x\right|\\&\leqslant\frac{2}{|x|}<\frac{2}{x'}.\end{aligned}$$

因而

$$\begin{aligned}\frac{1}{2\delta}\left|\int_{-\delta}^{\delta}\varphi_{n_k}(t)dt\right|&\leqslant\frac{1}{2\delta}\left[2\delta\int_{|x|\leqslant x'}dF_{n_k}(x)+\frac{2}{x'}\int_{|x|>x'}dF_{n_k}(x)\right]\\&\leqslant\alpha_k+\frac{1}{\delta x'}\end{aligned}$$

$$<\alpha+\frac{\varepsilon}{4}+\frac{1}{\delta\cdot\frac{4}{\delta\varepsilon}}$$

$$=\alpha+\frac{\varepsilon}{2}.$$

由于$\lim\limits_{k\to\infty}\varphi_{n_k}(t)=\varphi(t)$，因而

$$\frac{1}{2\delta}\left|\int_{-\delta}^{\delta}\varphi(t)\,dt\right|\leqslant\alpha+\frac{\varepsilon}{2}.$$

这个式子与(5.4.17)式矛盾．因此，$F(\infty)-F(-\infty)=1$，即得$F(\infty)=1$，$F(-\infty)=0$．

上面我们证明了$\{F_{n_k}(x)\}$弱收敛于分布函数$F(x)$，事实上$\{F_n(x)\}$亦弱收敛于同一分布函数$F(x)$．否则，必存在另一与$\{F_{n_k}(x)\}$不同的子序列$\{F_{n_k'}(x)\}$，它弱收敛于另一个分布函数$\tilde{F}(x)$(至少在某一连续点x_0上$\tilde{F}(x_0)\neq F(x_0)$)．然而用上述证明方法得到同一特征函数$\varphi(t)$，有

$$\varphi(t)=\int_{-\infty}^{\infty}e^{jtx}d\tilde{F}(x),$$

由特征函数唯一决定分布函数这一事实，必有$\tilde{F}(x)=F(x)$．■

有了定理5.4.1，我们可叙述并证明常用的中心极限定理．

三、中心极限定理

定义 5.4.2 设$\xi_n(n=1,2,\cdots)$为相互独立的随机变数序列，有有限的数学期望和方差：

$$E(\xi_k)=a_k,\ D(\xi_k)=\sigma_k^2\quad(k=1,2,\cdots).$$

令

$$\left.\begin{aligned}B_n^2&=\sum_{k=1}^{n}D(\xi_k),\\ \eta_n&=\sum_{k=1}^{n}\frac{\xi_k-a_k}{B_n},\end{aligned}\right\}\quad(n=1,2,\cdots),$$

若对于 $x \in R_1$ 一致地有

$$\lim_{n\to\infty} P\{\eta_n < x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy, \tag{5.4.18}$$

则称随机序列$\{\xi_n\}$服从中心极限定理.

中心极限定理早在18世纪由隶莫佛尔首先提出，但到现在内容已经十分丰富，我们不可能在这里一一介绍，只作初步叙述.

定理5.4.2 隶莫佛尔-拉普拉斯定理.

设 $\xi_n (n=1,2,\cdots)$为相互独立且具有相同两点分布的随机序列，且

$$P\{\xi_k=1\}=p,\ P\{\xi_k=0\}=q$$

$(k=1,2,\cdots)$. 其中 $q=1-p$, $0<p<1$. 则$\{\xi_n\}$服从中心极限定理.

证 由假设可得

$$E(\xi_k)=p,\ D(\xi_k)=pq,\ (k=1,2,\cdots),$$

$$B_n^2=\sum_{k=1}^{n} D(\xi_k)=npq.$$

令

$$\zeta_n=\sum_{k=1}^{n}\xi_k,$$

$$\eta_n=\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-p)=\frac{\zeta_n-np}{\sqrt{npq}},\ (n=1,2,\cdots).$$

依定理5.4.1，只须证当$n\to\infty$时η_n的特征函数 $\varphi_{\eta_n}(t)\to e^{-\frac{1}{2}t^2}$即可.

由于ζ_n的特征函数为 $\varphi_{\zeta_n}(t)=(q+pe^{jt})^n$，故

$$\varphi_{\eta_n}(t)=e^{-j\frac{np}{\sqrt{npq}}t}\left[q+pe^{j\frac{1}{\sqrt{npq}}t}\right]^n$$

$$=\left[qe^{-j\frac{p}{\sqrt{npq}}t}+pe^{j\frac{q}{\sqrt{npq}}t}\right]^n. \tag{5.4.19}$$

应用 e^{jx}在$x=0$的邻域的泰勒展式:

$$e^{jx}=\sum_{k=0}^{n}\frac{(jx)^k}{k!}+o(x^n),$$

得

$$qe^{-j\frac{t}{\sqrt{npq}}}+pe^{j\frac{t}{\sqrt{npq}}}$$

$$=(p+q)-(p+q)\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)$$

$$=1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right). \tag{5.4.20}$$

由(5.4.19)及(5.4.20)两式得到，当$n\to\infty$时

$$\varphi_{\eta_n}(t)=\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n\to e^{-\frac{1}{2}t^2}.\ \blacksquare$$

用同样的证明方法，可以得到更一般的结果如下：

定理5.4.3 设 $\xi_n(n=1, 2, \cdots)$ 为相互独立同分布的随机序列，且

$$E(\xi_k)=a,$$
$$D(\xi_k)=\sigma^2<\infty\ \ (\sigma^2\neq 0),$$
$$(k=1,2,\cdots).$$

则$\{\xi_n\}$服从中心极限定理.

对于任意 x_1, x_2，由(5.4.18)式并且仍用前面的符号，当n充分大时有，

$$P\{x_1\leqslant\zeta_n<x_2\}$$
$$=P\left\{\frac{x_1-na}{\sqrt{n}\,\sigma}\leqslant\eta_n<\frac{x_2-na}{\sqrt{n}\,\sigma}\right\}$$
$$\approx\Phi\left(\frac{x_2-na}{\sqrt{n}\,\sigma}\right)-\Phi\left(\frac{x_1-na}{\sqrt{n}\,\sigma}\right), \tag{5.4.21}$$

其中

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}y^2}\,dy.$$

上述两个定理有许多应用的例子，下面介绍几个有趣的例子.

例5.4.4 利用上述定理结果，说明例5.4.1关于钉板试验的现象.

仍用例5.4.1的符号. 显然，$\{\xi_n\}$满足上述定理的条件. 那里$E(\xi_k)=0, D(\xi_k)=\sigma=1(k=1,2,\cdots)$.

因此，由定理5.4.3有

$$\lim_{n\to\infty} P\left\{\frac{\xi_n}{\sqrt{n}}<x\right\}=\Phi(x).$$

因而当n充分大时

$$P\{x_1\leqslant\xi_n<x_2\}=P\left\{\frac{x_1}{\sqrt{n}}\leqslant\frac{\xi_n}{\sqrt{n}}<\frac{x_2}{\sqrt{n}}\right\}$$

$$\approx\Phi\left(\frac{x_2}{\sqrt{n}}\right)-\Phi\left(\frac{x_1}{\sqrt{n}}\right).$$

在例5.4.1中，$n=16$(即有16层钉子). 则有

$$P\{x_1\leqslant\xi_{16}<x_2\}\approx\Phi\left(\frac{x_2}{4}\right)-\Phi\left(\frac{x_1}{4}\right).$$

现考虑独立地投入100个小球.

查正态分布表得

$$P\{0\leqslant\xi_{16}<1\}\approx 0.0987,$$

因而，我们得到，100个小球中，大约有$100\times0.0987\approx10$个小球落在[0,1]这一格中. 根据正态分布的对称性，同样大约有10个小球

表 5.4.1

区　　间	近似概率	近似球数	区　　间	近似概率	近似球数
[-1,0]或[0,1]	0.0987	10	[-6,-5]或[5,6]	0.0388	4
[-2,-1]或[1,2]	0.0928	9	[-7,-6]或[6,7]	0.0268	3
[-3,-2]或[2,3]	0.0819	8	[-8,-7]或[7,8]	0.0173	2
[-4,-3]或[3,4]	0.068	7	[-9,-8]或[8,9]	0.0105	1
[-5,-4]或[4,5]	0.053	5	[-10,-9]或[9,10]	0.006	0或1

落在[-1,0]这一格中．同理，可得如表5.4.1的结果．

例 5.4.5 设某单位有200台电话机，每台电话机大约有5%的时间要使用外线通话．若每台电话机是否使用外线是相互独立的，问该单位总机至少需要安装多少条外线，才能以90%以上的概率保证每台电话机需要使用外线时不被占用．

解 把每一台电话机是否使用外线作为一次独立试验．用 $\{\xi_k=1\}$ 表示第 k 台电话机使用外线这一事件，根据题设，$P\{\xi_k=1\}=0.05$. 200台电话机通话则相当于作200次贝努里试验，即 $n=200$．这样，

$$\zeta_n=\sum_{k=1}^{n}\xi_k\ （显然\ \zeta_n\geqslant 0）$$

表示同时使用外线的电话机的总数．我们的问题是求：使得

$$P\{0\leqslant\zeta_n\leqslant k\}\geqslant 90\%$$

成立的最小的 k 值．

令

$$\eta_n=\frac{\zeta_n-np}{\sqrt{npq}},$$

其中 $p=P\{\zeta_k=1\}=0.05$，$q=1-p=0.95$．依题意，$\{\xi_k\}$ 满足定理5.4.2的条件，故当 n 充分大时

$$P\{0\leqslant\zeta_n\leqslant k\}=P\left\{\frac{-np}{\sqrt{npq}}\leqslant\eta_n\leqslant\frac{k-np}{\sqrt{npq}}\right\}$$

$$\approx\frac{1}{\sqrt{2\pi}}\int_{-\frac{np}{\sqrt{npq}}}^{\frac{k-np}{\sqrt{npq}}}e^{-\frac{1}{2}y^2}dy$$

$$=\Phi\left(\frac{k-np}{\sqrt{npq}}\right)-\Phi\left(\frac{-np}{\sqrt{npq}}\right).$$

由题设知

$$np=200\times 0.05=10,\quad npq=10\times 0.95=9.5,$$

因此，

$$P\{0\leqslant\zeta_n\leqslant k\}\approx\Phi\left(\frac{k-10}{\sqrt{9.5}}\right)-\Phi\left(-\frac{10}{\sqrt{9.5}}\right).$$

由于

$$\Phi\left(-\frac{10}{\sqrt{9.5}}\right)\approx 0,$$

故得

$$P\{0\leqslant\zeta_n\leqslant k\}\approx\Phi\left(\frac{k-10}{\sqrt{9.5}}\right).$$

问题现在化为求这样一个最小 k 值，使满足

$$\Phi\left(\frac{k-10}{\sqrt{9.5}}\right)\geqslant 90\%.$$

查正态分布表得，

$$\Phi(1.30)=0.9032,$$

因此可得

$$\frac{k-10}{\sqrt{9.5}}\geqslant 1.30,$$

即

$$k\geqslant 14.$$

即若某单位有14条外线，则可保证有90%以上的概率使每一台电话机需要使用外线时不被占用.

例 5.4.6 设 $\xi_1,\xi_2,\cdots,\xi_{48}$ 为相互独立且相同分布的随机变数，其密度函数为

$$f(x)=\begin{cases}1, & 当 0\leqslant x\leqslant 1;\\ 0, & 其它.\end{cases}$$

我们可计算得

$$E(\xi_k)=\frac{1}{2},$$

$$D(\xi_k)=\frac{1}{12}.$$

因此，$\{\xi_k\}$满足定理 5.4.3 条件，故有.

$$P\{\eta_{48}<x\}\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}y^2}dy,$$

其中

$$\eta_{48}=\sum_{k=1}^{48}\frac{1}{\sqrt{48D(\xi_k)}}[\xi_k-E(\xi_k)]$$

$$=\sum_{k=1}^{48}\frac{1}{2}\left[\xi_k-\frac{1}{2}\right].$$

现考虑

$$\bar{\xi}_{48}=\frac{1}{48}\sum_{k=1}^{48}\xi_k,$$

则有

$$P\{\bar{\xi}_{48}<0.4\}=P\left\{\frac{\bar{\xi}_{48}-E(\xi_k)}{\sqrt{\frac{1}{48}D(\xi_k)}}<\frac{0.4-0.5}{\sqrt{\frac{1}{48}\times\frac{1}{12}}}\right\}$$

$$\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-2.4}e^{-\frac{1}{2}y^2}dy$$

$$\approx 0.008.$$

上面结果表明：虽然每一ξ_k，在$[0,1]$上取其中的每一值的可能性是一样的，但其平均值$\frac{1}{48}\sum_{k=1}^{48}\xi_k$取值小于0.4的可能性是很小的，差不多(以概率为99.2%)可以认为只在$[0.4,1]$内取值.

例5.4.7 设随机序列$\xi_n(n=1,2,\cdots)$相互独立且服从同一哥西分布，其密度函数为

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2}.$$

考虑

$$\eta_n=\frac{1}{n}\sum_{k=1}^{n}\xi_k.$$

由例4.1.9知ξ_k的特征函数为

$$\varphi_{\xi_k}(t)=e^{-|t|},$$

因此，η_n的特征函数为

$$\varphi_{\eta_n}(t)=e^{-n\frac{|t|}{n}}=e^{-|t|}.$$

这结果告诉我们，对于任意 n，η_n仍服从哥西分布，因此，当 $n\to\infty$ 时，η_n的极限分布不是正态分布. 故随机变数 $\xi_n(n=1,2,\cdots)$的前 n 项平均数$\frac{1}{n}\sum_{k=1}^{n}\xi_k$ 的分布函数不能用正态分布来近似. 究其原因，虽然$\{\xi_n\}$相互独立且同分布，但它不满足定理5.4.3中的条件(大家知道哥西分布的数学期望与方差都不存在)，因此$\{\xi_n\}$不服从中心极限定理.

定理5.4.2及定理5.4.3，给出了关于随机序列$\{\xi_n\}$是独立同分布时的中心极限定理，有时我们还必须考虑不是同分布时的情形. 由于中心极限定理的各种类型研究很多，我们不可能在本教材中详细介绍，下面只对满足中心极限定理的林德贝格充分条件作一介绍.

定理5.4.4　林德贝格定理

设独立随机序列 $\xi_n(n=1,2,\cdots)$满足林德贝格条件：即若对任意$\varepsilon>0$，有

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x)=0, \qquad (5.4.22)$$

其中$F_k(x)$是 ξ_k的分布函数，$a_k=E(\xi_k)$，$\sigma_k^2=D(\xi_k)$，$B_n^2=\sum_{k=1}^{n}\sigma_k^2$.

则对 x 一致地有

$$\lim_{n\to\infty} P\left\{\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-a_k)<x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}y^2}dy.$$

先证下面三个引理.

引理5.4.3 令 $\xi_{nk}=\dfrac{\xi_k-a_k}{B_n}$, $F_{nk}(x)$和$\varphi_{nk}(t)$分别为 ξ_{nk}的分布函数和特征函数,则在定理5.4.4 的条件下,对任意$t\in[-T,T]$(其中$T>0$)一致地有

$$\lim_{n\to\infty}\max_{1\leqslant k\leqslant n}|\varphi_{nk}(t)-1|=0.$$

***证** 由假设得

$$E(\xi_{nk})=0,\ D(\xi_{nk})=\frac{1}{B_n^2}D(\xi_k)\quad(k=1,2,\cdots,n),$$

并且

$$\sum_{k=1}^{n}D(\xi_{nk})=\sum_{k=1}^{n}\frac{1}{B_n^2}D(\xi_k)=1.$$

因此

$$\begin{aligned}\varphi_{nk}(t)-1&=\int_{-\infty}^{\infty}e^{jtx}dF_{nk}(x)-\int_{-\infty}^{\infty}dF_{nk}(x)\\&\quad-\int_{-\infty}^{\infty}jtx\,dF_{nk}(x)\\&=\int_{-\infty}^{\infty}(e^{jtx}-1-jtx)dF_{nk}(x).\qquad(5.4.23)\end{aligned}$$

由于

$$\left|e^{jtx}-1-jtx\right|\leqslant\frac{t^2}{2}x^2,$$

所以

$$|\varphi_{nk}(t)-1| \leqslant \frac{t^2}{2}\int_{-\infty}^{\infty} x^2 dF_{nk}(x)$$

$$= \frac{t^2}{2}\int_{|x|>\varepsilon} x^2 dF_{nk}(x) + \frac{t^2}{2}\int_{|x|\leqslant\varepsilon} x^2 dF_{nk}(x). \qquad (5.4.24)$$

注意到 $\xi_{nk}=\dfrac{\xi_k-a_k}{B_n}$，由假设得到，当 n 充分大时，下式

$$\int_{|x|>\varepsilon} x^2 dF_{nk}(x) = \int_{\left|\frac{x-a_k}{B_n}\right|>\varepsilon} \left(\frac{x-a_k}{B_n}\right)^2 dF_k(x)$$

$$= \frac{1}{B_n^2}\int_{|x-a_k|>\varepsilon B_n} (x-a_k)^2 dF_k(x) < \varepsilon^2$$

对任意 k $(1\leqslant k\leqslant n)$ 成立。

另一方面

$$\int_{|x|\leqslant\varepsilon} x^2 dF_{nk}(x) \leqslant \varepsilon^2\int_{-\infty}^{\infty} dF_{nk}(x) = \varepsilon^2,$$

因此，由(5.4.24)式得到，对任意有限区间$[-T,T]$中的 t 及 k $(1\leqslant k\leqslant n)$，当$n\geqslant N$时有

$$|\varphi_{nk}(t)-1| < \varepsilon^2 T^2,$$

因而关于 $t\in[-T,T]$ 一致地有

$$\max_{1\leqslant k\leqslant n}|\varphi_{nk}(t)-1| < \varepsilon^2 T^2. \quad \blacksquare$$

引理5.4.4 在定理5.4.4的条件下，若令

$$R_n(t) = \sum_{k=1}^{n}\sum_{s=2}^{\infty}\frac{(-1)^{s-1}}{s}[\varphi_{nk}(t)-1]^s,$$

则关于$t\in[-T,T]$一致地有

$$R_n(t)\to 0 \quad (n\to\infty),$$

其中$\varphi_{nk}(t)$如引理5.4.3所述。

*证 因为

$$|R_n(t)| \leqslant \sum_{k=1}^{n}\sum_{s=2}^{\infty}\frac{1}{2}|\varphi_{nk}(t)-1|^s$$

$$=\frac{1}{2}\sum_{k=1}^{n}\frac{|\varphi_{nk}(t)-1|^2}{1-|\varphi_{nk}(t)-1|} \qquad (5.4.25)$$

由引理5.4.3知道，当n足够大时，对任意k $(k=1,2,\cdots,n)$以及$t\in[-T,T]$一致地有

$$|\varphi_{nk}(t)-1|<\frac{1}{2},$$

因此

$$|R_n(t)| \leqslant \frac{1}{2}\sum_{k=1}^{n}\frac{|\varphi_{nk}(t)-1|^2}{1-\frac{1}{2}}$$

$$=\sum_{k=1}^{n}|\varphi_{nk}(t)-1|^2$$

$$\leqslant \max_{1\leqslant k\leqslant n}|\varphi_{nk}(t)-1|\sum_{k=1}^{n}|\varphi_{nk}(t)-1|. \qquad (5.4.26)$$

注意到$E(\xi_{nk})=0$，$\sum\limits_{k=1}^{n}D(\xi_{nk})=1$及(5.4.24)式，可得

$$\sum_{k=1}^{n}|\varphi_{nk}(t)-1| \leqslant \frac{t^2}{2}\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2dF_{nk}(x)$$

$$=\frac{t^2}{2}\sum_{k=1}^{n}D(\xi_{nk})=\frac{t^2}{2}. \qquad (5.4.27)$$

由(5.4.26)及(5.4.27)两式并根据引理5.4.3，可得对$t\in[-T,T]$一致地有

$$|R_n(t)| \leqslant \frac{t^2}{2}\max_{1\leqslant k\leqslant n}|\varphi_{nk}(t)-1| \to 0 \quad (n\to\infty). ∎$$

引理5.4.5 在定理5.4.4的条件下，若令

$$r_n(t)=\frac{t^2}{2}+\sum_{k=1}^{n}\int_{-\infty}^{\infty}(e^{jtx}-1-jtx)dF_{nk}(x),$$

则关于$t\in[-T,T]$一致地有

$$\lim_{n\to\infty}r_n(t)=0.$$

*证 由题设可得

$$\sum_{k=1}^{n}\int_{-\infty}^{\infty}\frac{(jtx)^2}{2}dF_{nk}(x)=-\frac{t^2}{2}\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2dF_{nk}(x)$$

$$=-\frac{t^2}{2}.$$

因而

$$r_n(t)=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[e^{jtx}-1-jtx-\frac{(jtx)^2}{2}\right]dF_{nk}(x)$$

$$=\sum_{k=1}^{n}\int_{|x|<\varepsilon}\left[e^{jtx}-1-jtx-\frac{(jtx)^2}{2}\right]dF_{nk}(x)$$

$$+\sum_{k=1}^{n}\int_{|x|>\varepsilon}\left[e^{jtx}-1-jtx-\frac{(jtx)^2}{2}\right]dF_{nk}(x).$$

由于

$$|e^{jtx}-1-jtx|\leqslant\frac{t^2x^2}{2},$$

$$\left|e^{jtx}-1-jtx-\frac{(jtx)^2}{2}\right|\leqslant\frac{1}{6}|t^3x^3|,$$

因此

$$|r_n(t)|\leqslant\frac{|t|^3}{6}\sum_{k=1}^{n}\int_{|x|<\varepsilon}|x|^3dF_{nk}(x)$$

$$+\sum_{k=1}^{n}\int_{|x|>\varepsilon}\left[|e^{jtx}-1-jtx|+\left|\frac{t^2x^2}{2}\right|\right]dF_{nk}(x)$$

$$\leqslant \frac{|t|^3}{6}\sum_{k=1}^{n}\varepsilon\int_{|x|\leqslant\varepsilon}x^2dF_{nk}(x)+\sum_{k=1}^{n}\left[\frac{t^2}{2}\int_{|x|>\varepsilon}x^2dF_{nk}(x)\right.$$

$$\left.+\frac{t^2}{2}\int_{|x|>\varepsilon}x^2dF_{nk}(x)\right]$$

$$\leqslant \frac{|t|^3}{6}\varepsilon+t^2\sum_{k=1}^{n}\int_{|x|>\varepsilon}x^2dF_{nk}(x)$$

$$=\frac{|t|^3}{6}\varepsilon+t^2\sum_{k=1}^{n}\frac{1}{B_n^2}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x).$$

故对$|t|\leqslant T$

$$|r_n(t)|\leqslant\frac{T^3}{6}\varepsilon+T^2\sum_{k=1}^{n}\frac{1}{B_n^2}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x).$$

由ε的任意性及定理5.4.4的条件，当n充分大时，对$t\in[-T,T]$一致地可使得$|r_n(t)|$任意小. ∎

在上面证明引理5.4.3及引理5.4.5时，我们用了如下两个不等式

$$|e^{jtx}-1-jtx|\leqslant\frac{t^2x^2}{2},$$

$$\left|e^{jtx}-1-jtx-\frac{(jtx)^2}{2}\right|\leqslant\frac{1}{6}|t^3x^3|.$$

为了证明这两个不等式，我们只须证明成立如下不等式：

对任意实数α，有

$$|e^{j\alpha}-1|\leqslant|\alpha|;$$

$$|e^{j\alpha}-1-j\alpha|\leqslant\frac{1}{2!}\alpha^2;$$

$$\left|e^{j\alpha}-1-j\alpha+\frac{\alpha^2}{2}\right|\leqslant\frac{1}{3!}\alpha^3.$$

对于$\alpha=0$，上面三式显然成立. 设$\alpha>0$，则

$$|e^{j\alpha}-1|=\left|\int_0^\alpha e^{jx}dx\right|\leqslant\alpha;$$

$$|e^{j\alpha}-1-j\alpha|=\left|\int_0^\alpha(e^{jx}-1)dx\right|\leqslant\int_0^\alpha xdx=\frac{\alpha^2}{2!};$$

$$\left|e^{j\alpha}-1-j\alpha+\frac{\alpha^2}{2}\right|=\left|\int_0^\alpha(e^{jx}-1-jx)dx\right|$$

$$\leqslant \int_0^a |e^{jx}-1-jx|dx \leqslant \int_0^a \frac{x^2}{2!}dx = \frac{1}{3!}a^3.$$

由于上面三个不等式的两边都是 a 的偶函数，故它们对 $a<0$ 时亦成立.

下面我们转向证明定理5. 4. 4.

仍用引理5. 4. 3 的符号，则只须证明

$$\lim_{n\to\infty} P\left\{\sum_{k=1}^{n}\xi_{nk}<x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}dy.$$

如令 $\varphi_n(t)$ 为 $\sum_{k=1}^{n}\xi_{nk}$ 的特征函数，亦即需要证明在定理 5. 4. 4 的条件下

$$\lim_{n\to\infty}\varphi_n(t)=e^{-\frac{1}{2}t^2}.$$

或者

$$\lim_{n\to\infty}\ln\varphi_n(t)=-\frac{1}{2}t^2. \qquad (5.4.28)$$

下面我们来证明(5. 4. 28)式.

由引理5. 4. 3 知，当 n 充分大时，对任意 k $(1\leqslant k\leqslant n)$，有

$$|\varphi_{nk}(t)-1|<\frac{1}{2},$$

因此

$$\ln\varphi_n(t)=\sum_{k=1}^{n}\ln\varphi_{nk}(t)=\sum_{k=1}^{n}\ln[1+(\varphi_{nk}(t)-1)]$$

$$=\sum_{k=1}^{n}[\varphi_{nk}(t)-1]+R_n(t) \qquad (5.4.29)$$

(对数取主值)，其中，

$$R_n(t)=\sum_{k=1}^{n}\sum_{s=2}^{\infty}\frac{(-1)^{s-1}}{s}[\varphi_{nk}(t)-1]^s.$$

另一方面，注意到 $E(\xi_{nk})=0$，有

$$\sum_{k=1}^{n}[\varphi_{nk}(t)-1]=\sum_{k=1}^{n}\int_{-\infty}^{\infty}(e^{jtx}-1-jtx)dF_{nk}(x)$$
$$=-\frac{t^2}{2}+r_n(t), \tag{5.4.30}$$

其中

$$r_n(t)=\frac{t^2}{2}+\sum_{k=1}^{n}\int_{-\infty}^{\infty}(e^{jtx}-1-jtx)dF_{nk}(x).$$

因此，由(5.4.29)式及(5.4.30)式得到

$$\ln\varphi_n(t)=-\frac{t^2}{2}+R_n(t)+r_n(t).$$

由引理5.4.4及引理5.4.5立即可得(5.4.28)式成立.

现在我们来研究一下林德贝格条件的概率意义.对任意$\varepsilon>0$，让我们考虑事件$\left\{\max\limits_{1\leqslant k\leqslant n}\left|\frac{\xi_k-a_k}{B_n}\right|>\varepsilon\right\}$的概率. 有

$$P\left\{\max_{1\leqslant k\leqslant n}\left|\frac{\xi_k-a_k}{B_n}\right|>\varepsilon\right\}=P\left\{\max_{1\leqslant k\leqslant n}|\xi_k-a_k|>\varepsilon B_n\right\}$$
$$=P\left\{\bigcup_{k=1}^{n}[|\xi_k-a_k|>\varepsilon B_n]\right\}$$
$$\leqslant\sum_{k=1}^{n}P\{|\xi_k-a_k|>\varepsilon B_n\}$$
$$=\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}dF_k(x)$$
$$\leqslant\frac{1}{\varepsilon^2B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x).$$

因此，当林德贝格条件成立时，可得对任意小的ε，

$$\lim_{n\to\infty}P\left\{\max_{1\leqslant k\leqslant n}\left|\frac{\xi_k-a_k}{B_n}\right|>\varepsilon\right\}=0,$$

或等价地

$$\lim_{n\to\infty} P\left\{\max_{1\leqslant k\leqslant n}\left|\frac{\xi_k-a_k}{B_n}\right|\leqslant\varepsilon\right\}=1.$$

上式说明，$\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-a_k)$的每一被加项 $\frac{1}{B_n}(\xi_k-a_k)$，当n 充分大时，依概率一致地小. 这样，定理5.4.4 则说明，这样一些"影响一致地小"的随机变数之和的极限分布是正态分布.

推论（李亚普诺夫定理）

设 $\xi_n(n=1,2,\cdots)$ 为相互独立随机序列，若存在 $\delta>0$，使得

$$\lim_{n\to\infty}\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E|\xi_k-a_k|^{2+\delta}=0,$$

则$\{\xi_n\}$服从中心极限定理. 其中$a_k=E(\xi_k)$，$B_n^2=\sum_{k=1}^{n}D(\xi_k)$.

证 只须验证此时林德贝格条件成立即可. 事实上，当$n\to\infty$时，

$$\begin{aligned}&\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}(x-a_k)^2dF_k(x)\\&\leqslant\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\varepsilon B_n}\frac{|x-a_k|^{2+\delta}}{(\varepsilon B_n)^\delta}dF_k(x)\\&\leqslant\frac{1}{\varepsilon^\delta}\ \frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E|\xi_k-a_k|^{2+\delta}\to0.\end{aligned}$$

*§ 5.5 三种收敛的关系

定理5.5.1 若$\xi_n\xrightarrow{P}\xi$，则$\xi_n\xrightarrow{W}\xi$.

证 设$F_n(x)$及$F(x)$分别为ξ_n，ξ的分布函数，对任意实数x及$x'(x'<x)$由于

$$\{\xi<x'\}=\{\xi_n<x,\xi<x'\}\cup\{\xi_n\geqslant x,\xi<x'\}$$

$$\subset \{\xi_n < x\} \cup \{\xi_n \geqslant x, \xi < x'\},$$

因而有

$$F(x') \leqslant F_n(x) + P\{\xi_n \geqslant x, \xi < x'\}.$$

另一方面，由$\xi_n \xrightarrow{P} \xi$，则当$n \to \infty$时，有

$$P\{\xi_n \geqslant x, \xi < x'\} \leqslant P\{|\xi_n - \xi| \geqslant x - x'\} \to 0.$$

故得

$$F(x') \leqslant \lim_{n\to\infty} \inf F_n(x).$$

同理可证，对 $x'' > x$，有

$$F(x'') \geqslant \lim_{n\to\infty} \sup F_n(x).$$

所以，对于$x' < x < x''$，有

$$F(x') \leqslant \lim_{n\to\infty} \inf F_n(x) \leqslant \lim_{n\to\infty} \sup F_n(x) \leqslant F(x'').$$

若x是$F(x)$的连续点，令$x' \to x, x'' \to x$，则有

$$\lim_{n\to\infty} F_n(x) = F(x).$$

即$\xi_n \xrightarrow{W} \xi$. ∎

定理5.5.1的逆命题不真.

例5.5.1 设ξ_n，$(n=1,2,\cdots)$及ξ是相互独立相同分布的随机变数. 其分布列皆为

ξ_n	0	1
P_n	$\frac{1}{2}$	$\frac{1}{2}$

$(n=1,2,\cdots)$,

由于ξ_n与ξ有相同分布，显然$\xi_n \xrightarrow{W} \xi$. 然而，

$$P\left\{|\xi_n - \xi| > \frac{1}{2}\right\} = P\{\xi_n = 1, \xi = 0\}$$

$$+P\{\xi_n = 0, \xi = 1\} = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2},$$

则ξ_n不依概率收敛于ξ.

回忆§5.3，我们便得到这一章中所述的三个收敛的关系如下：

以概率为1收敛⟹依概率收敛⟹依分布收敛. 然而上述之逆均不真.

习题

1. 设$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy$. 试证明对每一$x>0$,有

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\left\{\frac{1}{x}-\frac{1}{x^3}\right\}<1-\Phi(x)<\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{x}.$$

2. 继上题. 试证明对每一常数a, 当$x\to\infty$时,有

$$\frac{\left\{1-\Phi\left(x+\frac{a}{x}\right)\right\}}{1-\Phi(x)}\longrightarrow e^{-a}.$$

3. 试证明：如果$\varphi(x)$是正的单调递增函数. 而$E[\varphi(|\xi|)]=m$存在. 则

$$P(|\xi|>t)\leqslant m/\varphi(t).$$

4. 设随机变数ξ服从指数—幂分布律：

$$f(x)=\frac{x^m}{m!}e^{-x}, x\geqslant 0.$$

试证明：$P\{0<\xi<2(m+1)\}>\frac{m}{m+1}$.

5. 设$\xi_n\xrightarrow{P}a$, $\eta_n\xrightarrow{P}b$ (a,b为常数且$b\neq 0$)，证明：1) $\xi_n^2\xrightarrow{P}a^2$；2) $\frac{\xi_n}{\eta_n}\xrightarrow{P}\frac{a}{b}$.

6. 若$\xi_n\xrightarrow{P}\xi$, 是否有$E(\xi_n)\to E(\xi)$？

7. 设$f(x)$在$(0,\infty)$上连续、单调上升. $f(0)=0$且$\sup\limits_{x\geqslant 0}f(x)<\infty$. 证明：$\xi_n\xrightarrow{P}0$等价于$\lim\limits_{n\to\infty}E[f(|\xi_n|)]=0$.

8. 设$\{\xi_k\}$ $(k=1,2,\cdots)$为随机序列，且

$$\xi_k \xrightarrow{W} c,$$

其中 c 为常数．则

$$\xi_k \xrightarrow{P} c.$$

9. 设$\{\xi_k\}$ $(k=1, 2,\cdots)$ 及 $\{\eta_k\}$ $(k=1, 2, \cdots)$ 是两个随机序列，且 $\xi_k \xrightarrow{W} \xi$, $\eta_k \xrightarrow{P} 0$. 证明：

$$\xi_k + \eta_k \xrightarrow{W} \xi.$$

10. 证明：如果$\xi_k \xrightarrow{P} \xi$ 且$\xi_k \xrightarrow{P} \eta$，则

$$P(\xi=\eta)=1.$$

11. 将 n 个带有号码 1 至 n 的球投入 n 个编有号码 1 至 n 的匣子．并限制每一个匣子只能放进一个球．设球与匣子的号码一致的个数是 s_n．试证：

$$\frac{s_n - E(s_n)}{n} \xrightarrow{P} 0.$$

12. 设 $g(x)$是 $-\infty<x<+\infty$ 上的连续函数，且 $\xi_n \xrightarrow{P} \xi$，则$g(\xi_n) \xrightarrow{P} g(\xi)$.

13. (马尔可夫大数定律) 设随机序列 $\{\xi_k\}$ 对任一正整数 n 均有 $D\left(\sum_{k=1}^{n}\xi_k\right)<\infty$，而且

$$\lim_{n\to\infty}\frac{1}{n^2} D\left(\sum_{k=1}^{n}\xi_k\right)=0,$$

则$\{\xi_k\}$服从大数定律．亦即，对任意 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E(\xi_k))\right|\geqslant\varepsilon\right\}=0.$$

14. 证明定理 5.1.3．并说明它是马尔可夫大数定律的特殊情形．而定理 5.1.2 又是定理 5.1.3 的特殊情形．

15. 证明：若$\{\xi_k\}$，$\{\eta_k\}$都服从大数定律，则$\{\xi_k \pm \eta_k\}$也服从大数定律．

16. 设 ξ_n 服从参数为 n, p_n 的二项分布，若当 $n\to\infty$ 时 $np_n\to\lambda$，用特征

函数法证明$\{\xi_n\}$的极限分布是参数为λ的泊松分布。

17. 用特征函数法证明辛钦大数定律.

18. 设随机序列$\{\xi_k\}$相互独立且满足

$$\xi_k=\begin{cases} k^{\lambda}, & \text{概率为}\dfrac{1}{2}, \\ -k^{\lambda}, & \text{概率为}\dfrac{1}{2}, \end{cases} \quad (k=1,2,\cdots),$$

证明：当$\lambda<\dfrac{1}{2}$时$\{\xi_k\}$服从大数定律.

19. 已知独立随机序列$\{\xi_n\}$具有同一分布函数：

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\operatorname{arc\,tg}\frac{x}{a},$$

试验证：辛钦大数定律对此序列是否适用.

20. 在每次试验中，事件A以概率$\dfrac{1}{2}$发生，是否可以用大于0.975的概率确信：在1000次试验中，事件A出现的次数在400与600范围内？

21. 试确定：由以下给定分布的相互独立的随机序列$\{\xi_n\}$是否满足使用大数定律的充分条件？

a) $P(\xi_k=\pm 2^k)=\dfrac{1}{2}$;

b) $P(\xi_k=\pm 2^k)=2^{-(2k+1)}$,
   $P(\xi_k=0)=1-2^{-2k}$;

c) $P(\xi_k=\pm k)=\dfrac{1}{2}k^{-\frac{1}{2}}$,

   $P(\xi_k=0)=1-k^{-\frac{1}{2}}$,

22. 随机序列$\{\xi_k\}$具有相同的期望与方差. 如果所有的协方差$b_{ij}=E(\xi_i-E(\xi_i))(\xi_j-E(\xi_j))<0(i\neq j)$，问大数定律对此序列是否适用？

23. 设有这样一个随机序列$\{\xi_n\}$，其ξ_n仅与ξ_{n-1}和ξ_{n+1}相关，而与其它所有的ξ_j不相关. 若$D(\xi_n)$一致有限，证明大数定律对此$\{\xi_k\}$成立.

24. 已知随机序列$\{\xi_k\}$的方差皆为有限，$D(\xi_k)\leqslant C$（C为常数）. 并且当$|i-j|\to\infty$时，$\rho_{ij}\to 0$.（ρ_{ij}为ξ_i与ξ_j的相关系数.）证明：大数定律对此序列适用.

25. 证明：如果对于独立随机序列$\{\xi_k\}$，当$A\to\infty$时，有

$$\max_{1\leq k\leq n}\int_{|x|\geq A}|x|dF_k(x)\to 0,$$

则在$\{\xi_k\}$上可应用大数定律.

26. 设$\{\xi_k\}$相互独立、同分布，其数学期望为$\mu(\xi_k\neq\mu)$，具有有限方差. 如果$s_n=\sum_{k=1}^{n}\xi_k$. 试证明：对序列$\{s_n\}$，大数定律不成立. 但如果$na_n\to 0$，则$\{a_ns_n\}$满足大数定律.

27. 设$\{\xi_k\}$为随机序列. $s_n=\sum_{k=1}^{n}\xi_k$. 若$|s_n|<Cn$，且$D(s_n)>an^2$ (C, a均为大于零的常数)，则大数定律不能应用于$\{\xi_k\}$.

28. 若$\{\xi_n\}$是单调下降取正值的随机序列，且$\xi_n\xrightarrow{P}0$，则$\xi_n\xrightarrow{a.s.}0$.

29. 设$\{\xi_n\}$为相互独立同分布的随机序列，令

$$\xi_N=\frac{1}{N}\sum_{k=1}^{N}[\xi_k-E(\xi_k)]^2.$$

证明$\xi_N\xrightarrow{a.s.}\sigma^2(\sigma^2=D(\xi_k))$.

30. 设$\{\xi_k\}$是一串相互独立的随机序列，且$\{\xi_k\}$服从参数为$\sqrt{k}$的泊松分布$\{P(j,\sqrt{k})\}$. 问$\{\xi_k\}$是否满足强大数定律?

31. 设$\{\xi_k\}$是相互独立的随机序列. 且$P(\xi_k=\pm 1)=\frac{1}{2}(1-2^{-k})$，而$P(\xi_k=\pm 2^k)=2^{-k-1}$. 证明：强、弱大数定律都可应用于$\{\xi_k\}$.

32. 设$\{\xi_k\}$为相互独立随机序列，证明：$P(\xi_k=\pm 2^k)=2^{-(k+1)}$，$P(\xi_k=0)=1-2^{-k}$服从强大数定律.

33. 设$\{\xi_k\}$为相互独立随机序列. ξ_k有分布密度

$$f_k(x)=\frac{1}{\sqrt[4]{k}\sqrt{\pi}}e^{-\frac{(x-\theta^k)^2}{\sqrt{k}}}$$

其中$0<\theta<1$. 问$\{\xi_k\}$是否服从强大数定律?

34. 证明：贝努里大数定律是隶莫佛尔-拉普拉斯极限定理的一个推论.

35. 证明：若$\{\xi_k\}$服从中心极限定理. 则$\{\xi_k+a_k\}$ (a_k为常数)也服从中心极限定理.

36. a) 证明：对独立随机序列$\{\xi_k\}$，如果$\lim_{n\to\infty}B_n=\infty$且$\lim_{n\to\infty}\sqrt{D\xi_n}$

$/B_n\to 0$，则中心极限定理成立的充分必要条件是林德贝格条件成立.

b) 第 21 题 b)的$\{\xi_k\}$是否服从中心极限定理？

37. 设$\{\xi_k\}$相互独立、并且 $P(\xi_k=\pm\sqrt{k})=\frac{1}{2}k^{-\frac{1}{3}}$，$P(\xi_k=\pm 1)=\frac{1}{2}(1-k^{-\frac{1}{3}})$. 问$\{\xi_k\}$是否满足中心极限定理？

38. 设$\{\xi_k\}$为相互独立的随机序列，ξ_k在$[-k,k]$上服从均匀分布. 问对$\{\xi_k\}$能否应用中心极限定理？

39. 设$\{\xi_k\}$为相互独立的随机序列，而且$\{\xi_k\}$一致有界. 即存在常数L，使对一切k，$P(|\xi_k|\leqslant L)=1$. 则当$B_n^2=\sum_{k=1}^{n}D(\xi_k)$趋于$\infty$时（当$n\to\infty$），中心极限定理成立.

40. 若独立随机序列$\{\xi_k\}$：

ξ_k取值	a_k	0	$-a_k$
ξ_k概率	p_k	$1-2p_k$	p_k

问a_k为何值时，大数定律及中心极限定理成立？

41. 对下列独立随机序列，李亚普诺夫定理是否成立？

i)

ξ_k取值	$-\sqrt{k}$	$\sqrt{k}$
ξ_k概率	$\frac{1}{2}$	$\frac{1}{2}$

ii)

ξ_k取值	$-k^a$	0	k^a
ξ_k概率	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$,a\geqslant 0$.

42. 证明：定理 5.4.2 是定理 5.4.4 的特例.

43. 设在第k次试验中，事件A的出现概率等于p_k，而S_n是事件A在n次独立试验中的出现次数. 求证：在$\sum_{k=1}^{\infty}p_kq_k=\infty$时（$q_k=1-p_k$），也只在这样的时候，有

$$P\left\{\frac{S_n - \sum_{k=1}^{n} \xi_k}{\sqrt{\sum_{k=1}^{n} p_k q_k}} < x\right\} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

附　录　I

排列组合补充

下面的结果大都可用归纳法证明，这里不赘述.

定理 1　设有 n_1 个字母 $a_{11},a_{12},\cdots a_{1n_1}$；$n_2$ 个字母 $a_{21},a_{22},\cdots a_{2n_2}$；…；$n_k$ 个字母 $a_{k1},a_{k2},\cdots,a_{kn_k}$. 现从 $a_{11},a_{12},\cdots a_{1n_1}$ 个任取一字母 a_{1i_1}；$a_{21},a_{22},\cdots a_{2n_2}$ 中任取一字母 a_{2i_2}；…；$a_{k1},a_{k2},\cdots a_{kn_k}$ 中任取一字母 a_{ki_k}. 则所得不同的字母列有 $n_1n_2\cdots n_k$ 个.

例 1　设有 n 件某种产品，每次任意抽取一件检查，检查完后放回，共抽取 r 次作为检查样品，求不同的样品列有多少？

解　此问题可视为定理 1 在 $n_1=n_2=\cdots n_k=n$ 时的情形. 因此，不同的样品列有 $\underbrace{n\times n\times\cdots\times n}_{r\text{个}}=n^r$ 个.

定理 2　设有 n_1 个字母 $a_{11},a_{12},\cdots a_{1n_1}$；$n_2$ 个字母 $a_{21},a_{22},\cdots a_{2n_2}$；…；$n_k$ 个字母 $a_{k_1},a_{k_2},\cdots,a_{kn_k}$. 如果于 $a_{11},a_{12},\cdots a_{1n_1}$ 中任取一个字母为 a_{1i_1} 后，$a_{21},a_{22},\cdots,a_{2n_2}$ 中只有 $n_2'(n_2'\leqslant n_2)$ 个可以取为下一个 a_{2i_2}，n_2' 为一常数，它与 a_{1i_1} 无关，但 $a_{21},a_{22},\cdots a_{2n_2}$ 中 n_2' 个可取为 a_{2i_2} 的字母可以与 a_{1i_1} 有关. 在 a_{1i_1}，a_{2i_2} 取定后，a_{31}，$a_{32},\cdots a_{3n_3}$ 只有某 n_3' 个字母可取为 a_{3i_3}，同样 n_3' 与 a_{1i_1}，a_{2i_2} 无关，但 a_{31},a_{32}，…，a_{3n_3} 中 n_3' 个字母可以取为 a_{3i_3} 的字母可以与 a_{1i_1}，a_{2i_2} 有关…. 最后，于 $a_{1i_1},a_{2i_2},\cdots,a_{k-1,i_{k-1}}$ 取定后，$a_{k1},a_{k2},\cdots,a_{kn_k}$ 中只有 n_k' 个字母允许取为 a_{ki_k}，n_k' 与 a_{1i_1},a_{2i_2}，…，$a_{k-1,i_{k-1}}$ 无关，但 a_{k1}，a_{k2}，$\cdots a_{kn_k}$ 中 n_k' 个可以取为 a_{ki_k} 的字母可以与 a_{1i_1}，

$a_{2i_2}, \cdots, a_{k-1, i_{k-1}}$ 有关，则在此情况下不同的字母列有 $n_1 \times n_2' \times n_3' \times \cdots \times n_k'$ 个.

推论 n 个字母 $a_1, a_2, \cdots, a_n$ 任取 r 个($r \leqslant n$)加以排列，则不同的排列共有$n(n-1)\cdots(n-r+1)=\frac{n!}{(n-r)!}$个. 特别当$r=n$时，得到 n 个字母的排列共有 $n(n-1)\cdots 2\cdot 1=n!$ 个.

令 $A_n^r=\frac{n!}{(n-r)!}$. 则我们常说“n 个字母任取 r 个字母的排列数为 A_n^r.”

定理 3 设有 n 个字母 $a_1, a_2, \cdots, a_n$，在其中任取 r 个，则由此 r 个字母组成不同的字母组有$\frac{n!}{r!(n-r)!}$种(这里两字母组称为相同，是指其构成组的成员，在不计及其次序的情况下，彼此全部相同).

证 设这种字母组数为 N，由于每一个数为 r 的字母组，可将其中字母依一切可能的顺序排列，由定理 2 之推论的特殊情形，可知共有 $r!$个字母列，因而每一字母组可产生 $r!$个字母列，由假设这种字母组有N个，则应产生 $N\cdot r!$个排列. 然由定理 2 之推论知 n 个字母任取 r 个字母的字母列有 $\frac{n!}{(n-r)!}$种. 故有

$$N\cdot r!=\frac{n!}{(n-r)!},$$

两边除以 $r!$，则最后得

$$N=\frac{n!}{r!(n-r)!}.$$

$\frac{n!}{r!(n-r)!}$在二项展开式中常用到，称为二项式系数，并以 C_n^r 表示，即 $C_n^r=\frac{n!}{r!(n-r)!}$. 这样，我们常说“$n$ 个字母任取 r 个的组合数”为 C_n^r 个.

例2 设有100件某产品，今欲抽取5件产品作样本检查，求

(1) 样品列共有多少个？

(2) 样品组共有多少个？

解 由定理2及定理3得样品列有$\frac{100!}{95!}$个，而样品组有$\frac{100!}{5!95!}$个.

定理4 n个物件分成k组，使第一组恰有n_1个；第二组恰有n_2个；…第k组恰有n_k个，$n_1, n_2, \cdots$为事先给定的非负整数. $n_1+n_2+\cdots+n_k=n$(两种分法认为相同，仅当两种分法的第一组与第一组相同，第二组与第二组相同，…第k组与第k组相同). 则不同的分法共有$\frac{n!}{n_1!n_2!\cdots n_k!}$种.

当$k=2$时，$n_1=r, n_2=n-r$，定理4即为定理3.

例3 有10件某产品分四个纸袋包装，第i个纸袋装i个($i=1,2,3,4$)问有多少种装法？

解 由定理4得，装法有

$$\frac{10!}{1!2!3!4!}=12600\text{ 种}.$$

定理5 n个物件，其中有n_1个带有号码"1"，n_2个带有号码"2"，…，n_k个带有号码"k"，且$n_1+n_2+\cdots+n_k=n$. 从这n个物件中任取r个物件，其中有r_i个带有号码"i"($i=1,2,\cdots k$)，$r_1+r_2+\cdots+r_k=r$且r_i为事先给定的非负整数(两次取法认为相同，仅当带有号码"i"的r_i个元素组成的元素组相同，$i=1,2,\cdots k$). 则不同的取法共有$C_{n_1}^{r_1}\cdot C_{n_2}^{r_2}\cdot\cdots\cdot C_{n_k}^{r_k}$种.

当$k=1$时，$n_1=n, r_1=r$，则定理5即为定理3.

例4 在52个某产品中，各有13个一等品，二等品，三等品，等外品. 今从52个产品中抽取13个，若要这13个中各有3个一

等品，二等品和三等品，而剩下的4个是等外品，问共有多少种取法？

解 按定理5，共有

$$C_{13}^{3}C_{13}^{3}C_{13}^{3}C_{13}^{4}=16726464040 \text{ 种取法.}$$

最后，我们介绍司蒂灵近似公式，在计算排列组合时，经常要计算 $n!$，虽然有 $\log n!$的表可查，但司蒂灵近似公式还是很有用。

司蒂灵近似公式为：

$$n!\approx\left(\frac{n}{e}\right)^{n}\sqrt{2n\pi},\qquad (n\geqslant 1),$$

或

$$e^{\frac{1}{12n}-\frac{1}{360n^{3}}}<\frac{n!}{\left(\frac{n}{e}\right)^{n}\sqrt{2n\pi}}<e^{\frac{1}{12n}},\qquad (n\geqslant 1).$$

附录 II

集合论简介

集合，常简称为集，它是集合论中最基本的概念，不能给予精确的数学定义。这里只给出一种非正式的描述。

我们把为了某种目的而研究的对象的全体称为集。例如：数 0 和 1；数 -1 和 1；全体自然数；全体实数；满足方程 $x^2-1=0$ 的正数；平面上所有的点；等等都是集。

集通常用 A, B, …或 Ω 等表示。今后讨论的集均认为是某一集的子集(见定义 1)，这个最大的集往往用 Ω 表示，Ω 称作空间。

集中的每一个对象称为元素或点。集中之元素常用 x，y，…a, b, …表示。若 x 是 A 中的元素就写为 $x\in A$(读作 x 属于 A)，如 x 不是集 A 的元素就写为 $x\overline{\in} A$(读作 x 不属于 A)。

例如：集 A 为全体自然数，则 $1\in A$ 而 $\frac{1}{2}\overline{\in} A$。

不含任何元素的集称为空集，常用 $\varnothing$ 表示。例如满足 $x^2+1=0$ 的实数的集是一空集。

有时集 A 中的元素可以全部写出，则可将集的元素写出后外加花括弧，例如集 $A=\{-1,1\}$。有时不能一一写出，或者即使可一一写出但为了简明、方便，可用文字或数学式子写出后外加花括弧，例如集 $A=\{$全体正整数$\}$。

元素、集、可用平面上的点来描述，它可以帮助我们理解集论中的定义和结果。

定义 1 设有 A, B 两集，若集 A 的所有元素都是集 B 的元素，

则称 A 是 B 的子集，写为 $A\subset B$，或 $B\supset A$. $A\subset B$（或$B\supset A$）又称为 A 包含于 B，或称 B 包含 A. 例如：$N=\{$全体自然数$\}$，$R=\{$全体实数$\}$，又如 $A=\{(x,y);\ x=y\}$，$S=\{$全体平面的点$\}$，则 $N\subset R$，$A\subset S$.

由定义，任意集 A 均有 $A\subset A$，$\varnothing\subset A$.

集的包含关系具有传递性，即若 $A\subset B$ 且 $B\subset C$，则 $A\subset C$，如图 1 所示.

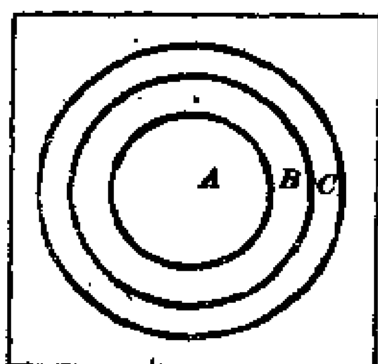


图 1

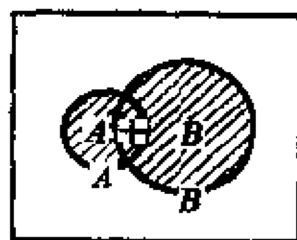


图 2

定义 2 设 A,B 为二集，若 $A\subset B$ 且 $B\subset A$，则称 A 与 B 相等，记为 $A=B$.

例如：$A=\{-1,-2\}$，$B=\{x: x^2+3x+2=0\}$，则有 $A=B$.

定义 3 设 A,B 为两个集，集 C 包含 A 与 B 中所有元素，且不包含其他元素，则称 C 是 A 与 B 的并（和）集，记为 $C=A\cup B$，或 $C=A+B$. 如图 2 所示.

例如：A 表示全体正有理数，B 表示全体正无理数，C 表示全体正实数，则有 $C=A\cup B$.

由定义 3，显然有 $A\cup B=B\cup A$.（交换律）. $(A\cup B)\cup C=A\cup(B\cup C)=A\cup B\cup C$.（结合律）. 且若 $A\subset B$，则 $A\cup B=B$.

显然有 $A\cup\varnothing=A$，$A\cup\Omega=\Omega$，$A\cup A=A$.

集合的并集可推广到有穷个或可列无穷个集合的情形，还可

以推广到不可列无穷集合的情形，我们并且把它记为 $C=\bigcup_{i\in I}A_i$，或 $C=\sum_{i\in I}A_i$，其中 I 为有穷、可列无穷或不可列无穷指标集．当 I 为有穷集时，记为 $C=\bigcup_{i=1}^{n}A_i$，或 $C=\sum_{i=1}^{n}A_i$．若 I 为可列无穷集时，记为 $C=\bigcup_{i=1}^{\infty}A_i$，或 $C=\sum_{i=1}^{\infty}A_i$．

定义 4 设 A，B 为两集，若集 D 包含 A 与 B 两集的所有共同的元素，且不包含任何其他元素，则称 D 为集 A 与集 B 的交或积，记为 $D=A\cap B$，或 $D=AB$．

例如：$A=\{1,2,3,4,5,6\}$，$B=\{4,5,6,7,8\}$，则 $D=\{4,5,6\}$．

集合的交或积可推广到有穷、可列无穷或不可列无穷集合的情形，并把它记为 $D=\bigcap_{i\in I}A_i$，或 $D=\prod_{i\in I}A_i$．I 可以是有穷，可列无穷或不可列无穷指标集．

由定义 4 可得

(1) $(A\cap B)\cap C=A\cap(B\cap C)=A\cap B\cap C$（结合律）．

(2) $A\cap A=A$;

$A\cap\varnothing=\varnothing$;

$A\cap\Omega=A$.

(3) 若 $A\subset B$，则 $A\cap B=A$．

由定义 3 及定义 4 可得

$A\cap(B\cup C)$

$=(A\cap B)\cup(A\cap C)$（乘法分配律）．

事实上，若

$x\in A\cap(B\cup C)\Rightarrow x\in A$，并且

$x\in(B\cup C)$，

$\Rightarrow x\in A$ 且 $x\in B$，或

$x\in A$ 且 $x\in C$，

$\Rightarrow x\in(A\cap B)\cup(A\cap C)$，即得

$$A\cap(B\cup C)\subset(A\cap B)\cup(A\cap C).$$

同理可得

$$A\cap(B\cup C)\supset(A\cap B)\cup(A\cap C).$$

因而

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C).$$

用同样方法可证得

$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$（加法分配律）。

定义 5 若 A,B 两集没有共同元素，即若

$$A\cap B=\varnothing,$$

则称集 A 与集 B 互斥或互不相交.

例如，$A=\{1,2,3,4\}$，$B=\{5,6,7,8\}$，则显然 A 与 B 互斥.

若集合 $A_1,A_2,\cdots$ 中任意两个集 A_i,A_j 都没有共同元素，则称 $A_1,A_2,\cdots$ 互斥或互不相交.

例如，$A_k=\left[\frac{1}{2^{k+1}},\frac{1}{2^k}\right)$，$k=1,2,\cdots$ 则 $A_1,A_2,\cdots$ 互斥.

定义 6 若 A,B 为两集，C 为包含属于 A 集而不属于 B 集的所有元素，但不包含别的元素，则称 C 为集 A 与集 B 的差集. 记为

$$C=A\setminus B,$$

或

$$C=A-B.$$

图 3 中阴影部分即表示 $A-B$.

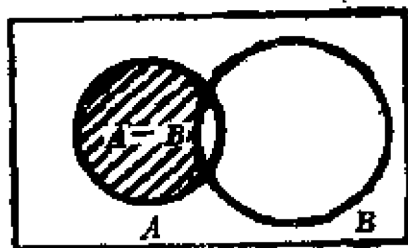


图 3

例如，设 A 为全体正实数，B 为全体正有理数，C 为全体正无理数，则 $C=A-B$.

特别，Ω 与 A 集的差集 $\Omega-A$，称为 A 的余集，或称为 A 集的逆．记为 $\bar{A}$，图 4 中阴影部分表示 $\bar{A}$.

由定义 6 可得

(1) $\overline{\varnothing}=\Omega$，$\overline{\Omega}=\varnothing$；

$\bar{\bar{A}}=A$, $A\cup\bar{A}=\Omega$；

$A\cap\bar{A}=\varnothing$；

(2) 若 $A\subset B$，则 $\bar{A}\supset\bar{B}$；

若 $A=B$，则 $\bar{A}=\bar{B}$；

(3) $A-B=A\cap\bar{B}=A-(A\cap B)$；

(4) $A-\varnothing=A$, $A-\Omega=\varnothing$；

(5) $A\cap(B-C)=(A\cap B)-(A\cap C)$.

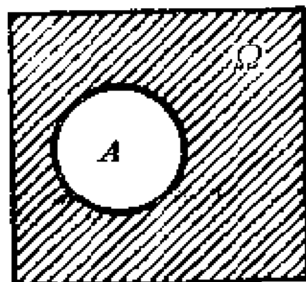


图 4

注意，$(A-B)\cup B$ 不一定等于 A，例如，当 A 与 B 满足下述关系时：

$$A\cap B=\varnothing,$$
$$B\neq\varnothing,$$

则有

$$(A-B)\cup B=A\cup B.$$

最后给出一个计算上很有用的公式，读者可以给予证明．

隶莫根公式：

$$\overline{(A\cup B)}=\bar{A}\cap\bar{B},$$
$$\overline{(A\cap B)}=\bar{A}\cup\bar{B}.$$

上述公式可推广到有穷，可列无穷或不可列无穷的情形：

$$\overline{\left(\bigcup_{i\in I}A_i\right)}=\bigcap_{i\in I}\bar{A}_i,$$
$$\overline{\left(\bigcap_{i\in I}A_i\right)}=\bigcup_{i\in I}\bar{A}_i.$$

附 录 III

R-S积分

下面简单介绍黎曼-斯蒂阶积分，不作详细证明.

§1 在有限闭区间上的一元 R-S 积分

限于本概率统计教程的需要，让我们引入较特殊的 R-S 积分定义如下：

定义 1.1 设 $f(x)$, $g(x)$为定义在$[a,b]$上取实值的函数. 其中 $g(x)$单调上升任作如下分点(或说作一剖分)：

$$a=x_0<x_1<x_2<\cdots<x_n=b,$$

并任取点 $x_k^*\in[x_k,x_{k+1}]$, $(k=0,1,\cdots,n-1)$.

作和式

$$\sigma=\sum_{k=0}^{n-1}f(x_k^*)[g(x_{k+1})-g(x_k)].$$

若存在实数 I，使对任给 $\varepsilon>0$，存在 $\delta>0$，只要

$$\lambda=\max_{0\leqslant k\leqslant n-1}(x_{k+1}-x_k)<\delta$$

时，不等式

$$|\sigma-I|<\varepsilon$$

对任意分点及任意的 x_k^* 取法皆成立，则记

$$(R)\int_a^b f(x)dg(x)=\lim_{\lambda\to 0}\sigma=\lim_{\lambda\to 0}\sum_{k=0}^{n-1}f(x_k^*)[g(x_{k+1})-g(x_k)]$$
$$=I,$$

并称 $I=(R)\int_a^b f(x)dg(x)$ 为 $f(x)$ 关于 $g(x)$ 在 $[a,\ b]$ 上的黎曼-斯蒂阶积分，简称 R-S 积分，并可简记为 $\int_a^b f(x)dg(x)$.

R-S 积分有下列易见的性质：

0) $\int_a^b f(x)dg(x)$ 存在的充要条件为：任给 $\varepsilon>0$，存在 $\delta>0$. 使对任意 $\lambda',\lambda''<\delta$ 时，有

$$|\sigma_{\lambda'}-\sigma_{\lambda''}|<\varepsilon.$$

(其中 σ_λ 表对满足 $\lambda=\max\limits_{0\leqslant k\leqslant n-1}(x_{k+1}-x_k)$ 的任一剖分所得的和式.)

1) $\int_a^b [f_1(x)\pm f_2(x)]dg(x)$

$$=\int_a^b f_1(x)dg(x)\pm\int_a^b f_2(x)dg(x).$$

2) $\int_a^b f(x)d[g_1(x)\pm g_2(x)]$

$$=\int_a^b f(x)dg_1(x)\pm\int_a^b f(x)dg_2(x).$$

3) 设 α,β 为任意二常数，则

$$\int_a^b \alpha f(x)d[\beta g(x)]=\alpha\beta\int_a^b f(x)dg(x).$$

上面三个等式成立的意义是：当等号右边存在时，左边也存在且相等.

4) 若 $a<c<b$，$\int_a^b f(x)dg(x)$ 存在，则两个积分 $\int_a^c f(x)dg(x)$ 与 $\int_c^b f(x)dg(x)$ 皆存在，且

$$\int_a^b f(x)dg(x)=\int_a^c f(x)dg(x)+\int_c^b f(x)dg(x)\text{①}.$$

---

① 注意：由 $\int_a^c f(x)dg(x)$ 与 $\int_c^b f(x)dg(x)$ 的存在性，并不能推出 $\int_a^b f(x)dg(x)$ 的存在. 而当 c 点为 $g(x)$ 之连续点时，则成立.

5) 若$\int_a^b f(x)dg(x)$与$\int_a^b g(x)df(x)$中有一个存在，则另一个也存在，且

$$\int_a^b f(x)dg(x)+\int_a^b g(x)df(x)=[f(x)g(x)]_a^b,$$

其中

$$[f(x)g(x)]_a^b=f(b)g(b)-f(a)g(a).$$

关于上述性质的证明，仅作提要如下：

1)，2)，3)，5)的证明，可直接由 R-S 积分定义证明。

4) 的证明可用性质0)或直接由定义证明.

下面给出一个 R-S 积分存在的条件：

定理 1.1 若 $f(x)$在$[a,b]$上连续，则

$$\int_a^b f(x)dg(x)$$

存在.

证 类似于普通黎曼积分的证明方法，通过达布大和与小和之差趋于 0 来证明. 只须注意，由于 $g(x)$单调，分点增加时，大和不增，小和不降这一性质.

关于 R-S 积分的计算，有下面的定理：

定理 1.2 若 $f(x)$在$[a,b]$上连续，$g(x)$的导数 $g'(x)$于$[a,b]$上存在且 $g'(x)$为黎曼可积，则$\int_a^b f(x)dg(x)$存在且

$$\int_a^b f(x)dg(x)=\int_a^b f(x)g'(x)dx.$$

证 首先，由于 $f(x)$在$[a,b]$连续，则必于$[a,b]$上黎曼可积. 从而 $f(x)g'(x)$于$[a,b]$上可积. 然后用拉格朗日中值公式，则可完成定理 1.2 的证明.

定理 1.3 若 $f(x)$在$[a,b]$上连续，设

$$a=c_0<c_1<c_2<\cdots<c_n=b,$$

若 $g(x)$在$[c_k,c_{k+1})$上取常数值，则$\int_a^b f(x)dg(x)$存在且

$$\int_a^b f(x)dg(x)=f(a)[g(a+0)-g(a)]$$
$$+\sum_{k=1}^{n-1}f(c_k)[g(c_k+0)-g(c_k-0)]+f(b)[g(b)-g(b-0)].$$

证 利用 R-S 积分性质 4) 即可得证.

§2 在无穷区间$(-\infty,\infty)$上的一元 R-S 积分

在本节中，恒设 $f(x)$, $g(x)$皆在任一闭区间$[a,b]$上有定义，其中 $g(x)$在$(-\infty,\infty)$上有界单调上升且$\int_a^b f(x)dg(x)$存在.

类似于普通积分的情形，引入如下定义：

定义 2.1 若

$$\lim_{\substack{a\to-\infty\\ b\to\infty}}\int_a^b f(x)dg(x)$$

存在，则记

$$I=(R)\int_{-\infty}^{\infty}f(x)dg(x),$$

并称 I 为 $f(x)$关于 $g(x)$在$(-\infty,\infty)$上的广义 R-S 积分.

以后为简单起见，$(R)\int_{-\infty}^{\infty}f(x)dg(x)$简记为$\int_{-\infty}^{\infty}f(x)dg(x)$.

容易把有限区间$[a,b]$上的 R-S 积分的性质 1)—5)全部推广到广义 R-S 积分①.

若我们引入符号

---

① 在性质 5)中，此处自然应该规定

$$\lim_{\substack{a\to-\infty\\ b\to\infty}}[f(x)g(x)]_a^b$$

存在(即极限为有限).

$$\chi_{[a,b]}(x)=\begin{cases}1, & \text{当 } x\in[a,b];\\ 0, & \text{其它},\end{cases}$$

则由定义显然可得

$$\int_a^b f(x)dg(x)=\int_{-\infty}^{\infty} f(x)\chi_{[a,b]}(x)dg(x).$$

对于广义 R-S 积分，也有类似于普通广义黎曼积分的性质:

1° $\int_{-\infty}^{\infty} f(x)dg(x)$ 存在的充分必要条件为 $\int_0^{\infty} f(x)dg(x)$ 与 $\int_{-\infty}^{0} f(x)dg(x)$ 存在.（其中，$\int_0^{\infty} f(x)dg(x)$ 定义为 $\lim\limits_{b\to\infty}\int_0^b f(x)dg(x)$，余类似.）

2°（哥西判别法）$\int_0^{\infty} f(x)dg(x)$ 存在的充分必要条件为：任给 $e>0$，存在 N，使当 $A', A''>N$ 时，

$$\left|\int_{A'}^{A''} f(x)dg(x)\right|<e.$$

3° 若 $\int_0^{\infty}|f(x)|dg(x)$ 存在，则必可得 $\int_0^{\infty} f(x)dg(x)$ 存在. 此时称积分 $\int_0^{\infty} f(x)dg(x)$ 为绝对收敛.

4° 设 $\int_0^{\infty}|f(x)|dg(x)$ 存在，则

$$\left|\int_0^{\infty} f(x)dg(x)\right|\leqslant\int_0^{\infty}|f(x)|dg(x).$$

以上四个性质的证明与普通黎曼积分一样.

5°（许瓦兹不等式）设 $f_1(x), f_2(x)$ 皆为平方可积. 即

$$\int_{-\infty}^{\infty} f_i^2(x)dg(x)<\infty \quad (i=1,2).$$

则

$$\int_{-\infty}^{\infty} f_1(x)f_2(x)dg(x)$$

存在且

$$\left(\int_{-\infty}^{\infty} f_1(x) f_2(x) dg(x)\right)^2 \leqslant \int_{-\infty}^{\infty} f_1^2(x) dg(x) \cdot \int_{-\infty}^{\infty} f_2^2(x) dg(x).$$

证 由设与3°及

$$|f_1 \cdot f_2| \leqslant \frac{1}{2}(|f_1|^2 + |f_2|^2)$$

可以推知$\int_{-\infty}^{+\infty} f_1(x) f_2(x) dg(x)$存在.现注意到

$$\begin{aligned} 0 &\leqslant \int_{-\infty}^{\infty} (f_1(x) - \lambda f_2(x))^2 dg(x) \\ &= \int_{-\infty}^{\infty} f_1^2(x) dg(x) - 2\lambda \int_{-\infty}^{\infty} f_1(x) f_2(x) dg(x) \\ &\quad + \lambda^2 \int_{-\infty}^{\infty} f_2^2(x) dg(x), \end{aligned}$$

由于λ的二次式非负,故得

$$\left(-2\int_{-\infty}^{\infty} f_1(x) f_2(x) dg(x)\right)^2 - 4\int_{-\infty}^{\infty} f_1^2(x) dg(x) \int_{-\infty}^{\infty} f_2^2(x) dg(x) \leqslant 0$$

由此得证5°. ▌

6° 设$\int_{-\infty}^{\infty} f_1(x) dg(x)$, $\int_{-\infty}^{\infty} f_2(x) dg(x)$皆存在,且对一切$x$, $f_1(x) \geqslant f_2(x)$,则有

$$\int_{-\infty}^{\infty} f_1(x) dg(x) \geqslant \int_{-\infty}^{\infty} f_2(x) dg(x).$$

下面叙述几个常用的关于广义R-S积分的定理.

定理 2.1 若$f(x)$在$(-\infty, \infty)$上连续且有界,则

$$\int_{-\infty}^{\infty} f(x) dg(x)$$

存在.

证 首先,由于$f(x)$在$(-\infty, \infty)$连续,由定理1.1得,对任意$a < b$,有

$$\int_a^b f(x)dg(x)$$

存在. 然后再用§2 性质 2°便可得定理 2.1 的证明 ∎

由于定理 2.1 中关于 $f(x)$为有界的假设太强，在概率论中常常不能满足. 因此，让我们来引入另一种条件.

称 $f(x)$满足条件(A)，若

1° $f(x)$在$(-\infty,\infty)$上连续，

2° $\int_{-\infty}^{\infty} f(x)dg(x)$ 存在.

定理 2.2 设 $f(x)$满足条件 (A)，并再假设 $g'(x)$ 对任意 $x\in(-\infty,\infty)$存在，且 $g'(x)$在任意有限区间$[a,b]$上黎曼可积. 则

$$\int_{-\infty}^{\infty} f(x)dg(x)=\int_{-\infty}^{\infty} f(x)g'(x)dx.$$

证 由定理 1.2 再通过取极限即得定理 2.2 的结果.

定理 2.3 设 $f(x)$ 满足条件 (A)，并再设存在实数序列 c_k，$k=0,\pm1,\pm2,\cdots$，使

$$\cdots<c_{-1}<c_0<c_1<\cdots$$

且 $g(x)$在$[c_k,c_{k+1})$上取常数值$(k=0,\pm1,\cdots)$，则

$$\int_{-\infty}^{\infty} f(x)dg(x)=\sum_{k=-\infty}^{\infty} f(c_k)[g(c_k+0)-g(c_k-0)].$$

证 由定理 1.3 再取极限即可得定理 2.3 的结果.

定理 2.4 设 $f_n(x)(n=1,2,\cdots)$，$f(x)$满足条件(A)，并且$f_n(x)$于$(-\infty,\infty)$一致收敛于 $f(x)$，则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dg(x)=\int_{-\infty}^{\infty} f(x)dg(x).$$

定理 2.5 设$f_n(x)$ $(n=1,2,\cdots)$，$f(x)$ 均满足条件(A)，且$f_n(x)$一致有界(即存在常数 k，使对 $n=1$，2，$\cdots$，均有$|f_n(x)|\leqslant k$)，$\lim\limits_{n\to\infty}f_n(x)=f(x)$，则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dg(x)=\int_{-\infty}^{\infty} f(x)dg(x).$$

上述定理 2.4 及定理 2.5 是关于在积分号下能够取极限的两个定理. 它们对于有限区间也成立. 这里证明均从略.

最后，我们介绍关于积分的变数代换问题:

定理 2.6 设 $f(x)$满足条件(A)，若 $x=x(t)$ 在 $t\in(-\infty,\infty)$ 上连续，单调上升且

$$\lim_{t\to\infty}x(t)=\infty,\qquad \lim_{t\to-\infty}x(t)=-\infty,$$

则

$$\int_{-\infty}^{\infty} f(x)dg(x)=\int_{-\infty}^{\infty} f[x(t)]dg[x(t)].$$

定理 2.7 设 $f(x)$满足条件(A)，且

$$g(x)=\int_c^x\varphi(x)dh(x),$$

其中 c 为常数，$\varphi(x)\geqslant 0$ 连续，$h(x)$有界上升，则

$$\int_{-\infty}^{\infty} f(x)dg(x)=\int_{-\infty}^{\infty} f(x)\varphi(x)dh(x).$$

§3 二维的 R-S积分

下面恒设二元函数 $F(x,y)$ 满足下述条件:

1° 对于平面上任意的矩形 $a_1\leqslant x\leqslant b_1, a_2\leqslant y\leqslant b_2$，有

$$F(b_1,b_2)-F(b_1,a_2)-F(a_1,b_2)+F(a_1,a_2)\geqslant 0$$

记上式左边为 $\Delta F(a_1,b_1;a_2,b_2)$.

2° $\lim\limits_{\substack{x\to\infty\\ y\to\infty}}F(x,y)=1$

且对任意 x 或 y，

$$\lim_{x\to-\infty}F(x,y)=\lim_{y\to-\infty}F(x,y)=0.$$

易知，由 1° 与 2° 可得 $F(x,y)$ 对每一变元单调. 亦即，当

$x_1 \leqslant x_2$，有 $F(x_1,y) \leqslant F(x_2,y)$，当 $y_1 \leqslant y_2$，有 $F(x,y_1) \leqslant F(x,y_2)$ 并且

$$\lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x,y) = 0.$$

现设 $f(x,y)$ 定义在整个平面上且取实值.

定义 3.1 在矩形 $a \leqslant x \leqslant b, c \leqslant y \leqslant d$ 上任作剖分：

$$a = x_0 < x_1 < \cdots < x_n = b,$$

$$c = y_0 < y_1 < \cdots < y_m = d.$$

任取点 $x_k \leqslant x_k^* \leqslant x_{k+1}, y_k \leqslant y_k^* \leqslant y_{k+1}$ 作和式

$$\sigma = \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} f(x_i^*, y_j^*) \cdot \Delta F(x_i, x_{i+1}; y_j, y_{j+1}).$$

若存在实数 I，使得对于任意给定的 $\varepsilon > 0$，存在 $\delta > 0$，只要 $\lambda = \max\limits_{\substack{0 \leqslant i \leqslant n-1 \\ 0 \leqslant j \leqslant m-1}} \{(x_{i+1} - x_i), (y_{j+1} - y_j)\} < \delta$ 时有

$$|\sigma - I| < \varepsilon,$$

则记

$$\iint\limits_{\substack{a \leqslant x \leqslant b \\ c \leqslant y \leqslant d}} f(x,y)\, dF(x,y) = \lim_{\lambda \to 0} \sigma$$

$$= \lim_{\lambda \to 0} \sum_{j=0}^{m-1} \sum_{i=0}^{n-1} f(x_i^*, y_j^*) \cdot \Delta F(x_i, x_{i+1}; y_j, y_{j+1}),$$

并称

$$I = (R) \iint\limits_{\substack{a \leqslant x \leqslant b \\ c \leqslant y \leqslant d}} f(x,y)\, dF(x,y)$$

为 $f(x,y)$ 关于 $F(x,y)$ 在矩形 $\{(x,y): a \leqslant x \leqslant b, c \leqslant y \leqslant d\}$ 上的 R-S 积分. 在不引起混淆的情形下，常常把前面的符号 (R) 省略.

定义 3.2 若

$$\lim_{\substack{b, d \to \infty \\ a, c \to -\infty}} (R) \iint\limits_{\substack{a \leqslant x \leqslant b \\ c \leqslant y \leqslant d}} f(x,y)\, dF(x,y)$$

存在，记为 I，则说 $f(x,y)$ 关于 $F(x,y)$ 在整个平面上有 R-S 积分存在，并称 I 为其积分值. 记作

$$I=(R)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dF(x,y).$$

易知，对于平面上的 R-S 积分，同样地可以有 §1 的性质1)—4) 成立①. 并且相应于 §2 的性质 1°—6° 亦成立. (性质 1° 自然地相应改为：

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dF(x,y)\text{ 存在}\Longleftrightarrow$$

$$\iint\limits_{\substack{0\leqslant x<\infty\\0\leqslant y<\infty}}f(x,y)\,dF(x,y),$$

$$\iint\limits_{\substack{0\leqslant x<\infty\\-\infty<y\leqslant 0}}f(x,y)\,dF(x,y),$$

$$\iint\limits_{\substack{-\infty<x\leqslant 0\\0\leqslant y<\infty}}f(x,y)\,dF(x,y),$$

及

$$\iint\limits_{\substack{-\infty<x\leqslant 0\\-\infty<y\leqslant 0}}f(x,y)\,dF(x,y)$$

存在.)

类似于一元情形，我们可以有下面的定理.

定理 3.1 若 $f(x,y)$ 在整个平面上连续有界，则

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dF(x,y)$$

---

①性质 4)相应地是把一个矩形分割成两个矩形或四个矩形，而在每个矩形上考虑二维积分即可. 例如：若 $a<e<b$，$c<f<d$，则矩形 $[a,b;\ c,d]$ 可分成 $[a,e;\ c,f]$，$[e,b;c,f]$，$[a,e;f,d]$ 及 $[e,b;f,d]$，于是

$$\iint\limits_{\substack{a\leqslant x\leqslant b\\c\leqslant y\leqslant d}}\cdots=\iint\limits_{\substack{a\leqslant x\leqslant e\\c\leqslant y\leqslant f}}\cdots+\iint\limits_{\substack{e\leqslant x\leqslant b\\c\leqslant y\leqslant f}}\cdots+\iint\limits_{\substack{a\leqslant x\leqslant e\\f\leqslant y\leqslant d}}\cdots+\iint\limits_{\substack{e\leqslant x\leqslant b\\f\leqslant y\leqslant d}}\cdots$$

存在.

与一维情形一样，定理3.1中关于$f(x,y)$为有界的假设太强. 故引入另一种条件.

称$f(x,y)$满足条件(B)，若

1° $f(x,y)$在整个平面上连续，

2° $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dF(x,y)$存在.

类似于一维情形，我们也可以有下列的定理.

定理 3.2 设$f(x,y)$满足条件(B)，且对任意x，$y\in(-\infty,\infty)$，有连续的二阶偏导数$\frac{\partial^2 F}{\partial x\partial y}(x,y)$存在，则

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dF(x,y)$$
$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\frac{\partial^2 F}{\partial x\partial y}(x,y)\,dx\,dy.$$

定理 3.3 若$f_n(x,y)$，$f(x,y)$满足条件(B)，且

$$f_n(x,y)\rightrightarrows f(x,y)$$

(一致收敛)，则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_n(x,y)dF(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dF(x,y).$$

定理 3.4 设$f_n(x,y)\ (n=1,2,\cdots)$，$f(x,y)$满足条件(B)，且$f_n(x,y)$一致有界，即对$x,y\in(-\infty,\infty)$

$$|f_n(x,y)|\leqslant K,$$

且

$$\lim_{n\to\infty} f_n(x,y)=f(x,y),$$

则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_n(x,y)dF(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dF(x,y).$$

现设存在单调不降函数$F_i(x)\ (i=1,2)$，满足

$$\left.\begin{array}{l}\lim\limits_{x\to+\infty}F_i(x)=1\\ \lim\limits_{x\to-\infty}F_i(x)=0\end{array}\right\}(i=1,2)$$

使
$$F(x,y)=F_1(x)\cdot F_2(y).\tag{3.1}$$

此时，若 $a_1\leqslant x\leqslant b_1,a_2\leqslant y\leqslant b_2$，则

$$\Delta F(a_1,b_1;a_2,b_2)$$
$$=[F_1(b_1)-F_1(a_1)]\cdot[F_2(b_2)-F_2(a_2)].\tag{3.2}$$

下面在(3.1)的假设下进行讨论

定理 3.5 设 $f(x,y)$ 满足条件(B)，且

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|f(x,y)|dF(x,y)$$

收敛，则

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dF(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dF_1(x)dF_2(y)$$
$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}f(x,y)dF_1(x)\right\}dF_2(y)$$
$$=\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}f(x,y)dF_2(y)\right\}dF_1(x).$$

这里第一个等号可以看作是由于有(3.2)式，而对二元积分规定的另一种记号．最后两个积分是累次积分．

现在稍微提一下上述诸定理的证明．定理 3.1—3.3 的证明类似于以前．定理 3.4 的证明需用到较复杂的工具．定理 3.5 可先对有限的矩形证明，然后再利用数学分析中关于二元极限与累次极限的关系使定理得到证明．

下面举例说明定理 3.5 在相互独立随机变数有关问题中的应用．

例 3.1 若 ξ_1,ξ_2 相互独立．设 $\xi_1+\xi_2$ 的分布函数为 $F_{\xi_1+\xi_2}(x)$，ξ_i 的分布函数为 $F_i(x)(i=1,2)$．ξ_1，ξ_2 的联合分布函数为 $F(x,y)$．则

$$F_{\xi_1+\xi_2}(x)=\int_{-\infty}^{\infty}F_1(x-z)dF_2(z)=\int_{-\infty}^{\infty}F_2(x-z)dF_1(z);$$

若 $\xi_2\neq 0$，则 ξ_1/ξ_2 的分布函数为

$$F_{\frac{\xi_1}{\xi_2}}(x)=\int_0^{\infty}F_1(xz)dF_2(z)+\int_{-\infty}^{0}[1-F_1(xz)]dF_2(z).$$

证 $F_{\xi_1+\xi_2}(x)=P\{\xi_1+\xi_2<x\}=\iint\limits_{x_1+x_2<x}dF(x_1,x_2)$

$$=\iint\limits_{x_1+x_2<x}dF_1(x_1)dF_2(x_2).$$

令

$$\begin{cases}z=x_1+x_2,\\ x_1=x_1.\end{cases}$$

则 $F_{\xi_1+\xi_2}=\iint\limits_{\substack{z<x\\-\infty<x_1<\infty}}dF_2(z-x_1)dF_1(x_1)=\int_{-\infty}^{\infty}dF_1(x_1)\int\limits_{-\infty<z<x}dF_2(z-x_1)$

$$=\int_{-\infty}^{\infty}F_2(x-x_1)dF_1(x_1).$$

第一式得证

$$F_{\frac{\xi_1}{\xi_2}}(x)=P\left\{\frac{\xi_1}{\xi_2}<x\right\}=\iint\limits_{\substack{\frac{x_1}{x_2}<x\\x_2\neq 0}}dF(x_1,x_2)=\iint\limits_{\substack{\frac{x_1}{x_2}<x\\x_2\neq 0}}dF_1(x_1)dF_2(x_2).$$

令

$$\begin{cases}z=\dfrac{x_1}{x_2},\quad (x_2\neq 0)\\ x_2=x_2.\end{cases}$$

则

$$\text{上式}=\iint\limits_{\substack{z<x\\-\infty<x_2<\infty}}dF_1(x_2z)dF_2(x_2)=\int_{-\infty}^{\infty}dF_2(x_2)\int\limits_{-\infty<z<x}dF_1(x_2z)$$

$$=\int_0^{\infty}dF_2(x_2)\int\limits_{-\infty<z<x}dF_1(x_2z)+\int_{-\infty}^{0}dF_2(x_2)\int\limits_{-\infty<z<x}dF_1(x_2z)$$

$$=\int_0^{\infty}F_1(xx_2)dF_2(x_2)+\int_{-\infty}^{0}[1-F_1(xx_2)]dF_2(x_2)$$

第二式得证. ▮

这里变量替换之可行，是由于在积分的和式中直接进行替换

后仍相等，因而取极限后（即它们的积分）相等。

§4 n 维的 R-S 积分

关于 n 维的 R-S 积分，同样可以用类似于定义平面上的 R-S 积分的方法作出. 但在此情形需假设 n 元函数 $F(x_1, x_2, \cdots, x_n)$ 满足：

1° 对于 n 维空间中任意的矩形 $a_1 \leqslant x_1 \leqslant b_1, \cdots, a_n \leqslant x_n \leqslant b_n$，有

$$\begin{aligned}&\Delta F(a_1, b_1; \cdots; a_n, b_n)\\&\equiv F(b_1, b_2, \cdots, b_n) - \sum_{i=1}^{n} P_i + \sum_{i<j} P_{ij} + \cdots\\&\qquad + (-1)^n F(a_1, a_2, \cdots, a_n) \geqslant 0;\end{aligned}$$

其中 $P_{i,j,\cdots,k}$ 表示函数 $F(x_1, x_2, \cdots, x_n)$ 在 $x_i = a_i, x_j = a_j, \cdots, x_k = a_k$ 而其余的 x_m 取 b_m 时的值.

2° $\lim\limits_{\substack{x_i \to \infty \\ 1 \leqslant i \leqslant n}} F(x_1, x_2, \cdots, x_n) = 1,$

且
$$\lim_{x_i \to -\infty} F(x_1, x_2, \cdots, x_n) = 0,$$

$$i = 1, 2, \cdots, n.$$

在此情形，R-S 积分

$$(R) \int \cdots \int_{\substack{a_1 \leqslant x_1 \leqslant b_1 \\ \cdots\cdots \\ a_n \leqslant x_n \leqslant b_n}} f(x_1, \cdots, x_n) dF(x_1, \cdots, x_n)$$

与

$$(R) \underbrace{\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{\infty}}_{n\text{ 重}} f(x_1, \cdots, x_n) dF(x_1, \cdots, x_n)$$

的定义，同样可以像二维情形那样，分别类似地作出. 并且，二维 R-S 积分之所有性质，也同样可以类似且平行地推广至 n 维 R-S 积分情形，也有类似性质成立。

表1 二项分布

$$b(n,k,p)=\sum_{i=k}^{n}\binom{n}{i}p^{i}(1-p)^{n-i}$$

n	k \ p	0.01	0.02	0.04	0.06	0.08	0.1	0.2	0.3	0.4	0.5	p / k	n
5	5			0.00000	0.00000	0.00000	0.00001	0.00032	0.00243	0.01024	0.03125	5	5
	4	0.00000	0.00000	00001	00006	00019	00046	00672	03078	08704	18750	4	
	3	00001	00008	00060	00197	00453	00856	05792	16308	31744	50000	3	
	2	00098	00384	01476	03187	05436	08146	26272	47178	66304	81250	2	
	1	04901	09608	18463	26610	34092	40951	67232	83193	92224	96875	1	
10	10								0.00001	0.00010	0.00098	10	10
	9							0.00000	00014	00168	01074	9	
	8						0.00000	00008	00159	01229	05469	8	
	7				0.00000	0.00000	00001	00086	01059	05476	17188	7	
	6			0.00000	00001	00004	00015	00637	04735	16624	37695	6	
	5		0.00000	00002	00015	00059	00163	03279	15027	36690	62305	5	
	4	0.00000	00003	00044	00203	00580	01280	12087	35039	61772	82813	4	
	3	00011	00086	00621	01884	04008	07019	32220	61722	83271	94531	3	
	2	00427	01618	05815	11759	18788	26390	62419	85069	95364	98926	2	
	1	09562	18293	33517	46138	56561	65132	89263	97175	99395	99902	1	
	15									0.00000	0.00003	15	
	14								0.00000	00003	00049	14	
	13								00001	00028	00369	13	
	12							0.00000	00009	00193	01758	12	
	11							00001	00067	00935	05923	11	

15	10							0.00011	0.00365	0.03383	0.15088	10	15
	9					0.00000	0.00000	00079	01524	09505	30362	9	
	8				0.00000	00001	00003	00424	05001	21310	50000	8	
	7			0.00000	00001	00008	00031	01806	13114	39019	69638	7	
	6		0.00000	00001	00015	00070	00225	06105	27838	59678	84912	6	
	5	0.00000	00001	00022	00140	00497	01272	16423	48451	78272	94077	5	
	4	00001	00018	00245	01036	02731	05556	35184	70313	90950	98242	4	
	3	00042	00304	02029	05713	11297	18406	60198	87317	97289	99631	3	
	2	00963	03634	11911	22624	34027	45096	83287	96473	99483	99951	2	
	1	13994	26143	45791	60471	71370	79411	96482	99525	99953	99997	1	
20	20										0.00000	20	20
	19									0.00000	00002	19	
	18									00001	00020	18	
	17								0.00000	00005	00129	17	
	16								00001	00032	00591	16	
	15								00004	00161	02069	15	
	14							0.00000	00026	00647	05766	14	
	13							00002	00128	02103	13159	13	
	12							00010	00514	06653	25172	12	
	11						0.00000	00056	01714	12752	41190	11	
	10					0.00000	00001	00258	04796	24466	58810	10	
	9				0.00000	00001	00006	00998	11333	40440	74828	9	
	8			0.00000	00001	00009	00042	03214	22773	58411	86841	8	
	7			00001	00011	00064	00239	08669	39199	74999	94234	7	
	6		0.00000	00010	00087	00380	01125	19579	58363	87440	97931	6	
	5	0.00000	00004	00096	00563	01834	04317	37035	76249	94905	99409	5	
	4	00004	00080	00741	02897	07862	13295	58855	89291	98404	99871	4	
	3	00100	00707	04386	11497	21806	32307	79392	96452	99639	99980	3	
	2	01686	05990	18966	33965	48314	60825	93082	99236	99948	99998	2	
	1	18209	35239	55800	70989	81131	87842	98847	99920	99996	1.00000	1	

表2 泊松(Poisson)分布

$$1-F(c)=\sum_{k=c}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}$$

c \ λ	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.010
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.0009995	0.0019980	0.0029955	0.0039920	0.0049875	0.0059820	0.0069756	0.0079681	0.0089596	0.0099502
2	0000005	0000020	0000045	0000080	0000125	0000179	0000244	0000318	0000403	0000497
3							0000001	0000001	0000001	0000002

c \ λ	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.0198013	0.0295545	0.0392106	0.0487706	0.0582355	0.0676062	0.0768837	0.0860688	0.0951626	0.1041659
2	0001973	0004411	0007790	0012091	0017296	0023386	0030343	0038150	0046788	0056241
3	0000013	0000044	0000104	0000201	0000344	0000542	0000804	0001136	0001547	0002043
4			0000001	0000003	0000005	0000008	0000016	0000025	0000033	0000056
5										0000001

c \ λ	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.21
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.1130796	0.1219046	0.1306418	0.1392920	0.1478562	0.1563352	0.1647298	0.1730409	0.1812692	0.1894158
2	0066491	0077522	0089316	0101858	0115132	0129122	0143812	0159187	0175231	0191931
3	0002633	0003323	0004119	0005029	0006058	0007212	0008498	0009920	0011485	0013197
4	0000079	0000107	0000143	0000187	0000240	0000304	0000379	0000467	0000563	0000685
5	0000002	0000003	0000004	0000006	0000008	0000010	0000014	0000018	0000023	0000029
6								0000001	0000001	0000001

c \ λ	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.40
0	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
1	0.1974812	0.2054664	0.2133721	0.2211992	0.2289484	0.2366205	0.2442163	0.2517364	0.2591818	0.3296800
2	0209271	0227237	0245815	0264990	0284750	0305080	0325968	0347400	0369363	0615519
3	0015060	0017083	0019266	0021615	0024135	0026829	0029701	0032755	0035995	0079263
4	0000819	0000971	0001142	0001334	0001548	0001786	0002049	0002339	0002658	0007763
5	0000036	0000044	0000054	0000066	0000080	0000096	0000113	0000134	0000158	0000612
6	0000001	0000002	0000002	0000003	0000003	0000004	0000005	0000006	0000008	0000040
7										0000002

c \ λ	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.393469	0.451188	0.503415	0.550671	0.593430	0.632121	0.667129	0.698806	0.727468	0.753403
2	090204	121901	155805	191208	227518	264241	300971	337373	373177	408167
3	014388	023115	034142	047423	062857	080301	099584	120513	142888	166502
4	001752	003358	005753	009080	013459	018988	025742	033769	043095	053725
5	000172	000394	000786	001411	002344	003660	005435	007746	010663	014253
6	000014	000039	000090	000184	000343	000594	000968	001500	002231	003201
7	000001	000003	000009	000021	000043	000083	000149	000251	000404	000622
8			000001	000002	000005	000010	000020	000037	000064	000107
9						000001	000002	000005	000009	000016
10								000001	000001	000002

c \ λ	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.776870	0.798103	0.817316	0.834701	0.850431	0.864665	0.877544	889197	899741	0.909282
2	442175	475069	506754	537163	566251	593994	620385	645430	669146	691559
3	191153	216642	242777	269379	296280	323324	350369	377286	403961	430291
4	065642	078813	093189	108708	125298	142877	161357	180648	200653	221277

续表

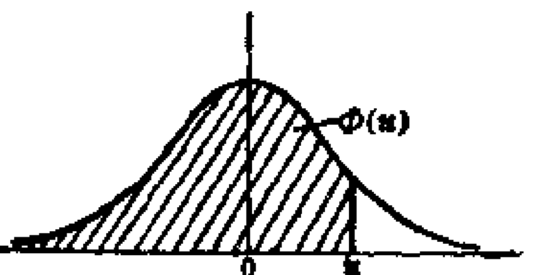
c \ λ	1.5	1.6	1.7	1.8	1.9	2.0	2.1	2.2	2.3	2.4
5	0.018576	0.023682	0.029615	0.036407	0.044081	0.052653	0.062126	0.072496	0.083751	0.095869
6	004456	006040	007999	010378	013219	016564	020449	024910	029976	035873
7	000926	001336	001875	002569	003446	004534	005862	007461	009362	011594
8	000170	000260	000388	000562	000793	001097	001486	001978	002589	003339
9	000028	000045	000072	000110	000163	000237	000337	000470	000642	000862
10	000004	000007	000012	000019	000030	000046	000069	000101	000144	000202
11	000001	000001	000002	000008	000005	000008	000013	000020	000029	000043
12					000001	000001	000002	000004	000006	000008
13								000001	000001	000002

c \ λ	2.5	2.6	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.4
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.917915	0.925726	0.932794	0.939190	0.944977	0.950213	0.954951	0.959238	0.963117	0.966627
2	712703	732615	751340	768922	785409	800852	815298	828799	841402	853158
3	456187	481570	506376	530546	554037	576810	598837	620096	640574	660260
4	242424	263998	285908	308063	330377	352768	375160	397480	419662	441643
5	108822	122577	137092	152324	168223	184737	201811	219387	237410	255818
6	042021	049037	056732	065110	074174	083918	094334	105408	117123	129458
7	014187	017170	020569	024411	028717	033509	038804	044619	050966	057853
8	004247	005334	006621	008131	009885	011905	014213	016830	019777	023074
9	001140	001487	001914	002433	003058	003803	004683	005714	006912	008293
10	000277	000376	000501	000660	000858	001102	001401	001762	002195	002709
11	000062	000087	000120	000164	000220	000292	000383	000497	000638	000810
12	000013	000018	000026	000037	000052	000071	000097	000129	000171	000223
13	000002	000004	000005	000008	000011	000016	000023	000031	000042	000057
14		000001	000001	000002	000002	000003	000005	000007	000010	000014
15						000001	000001	000001	000002	000003
16										000001

c \ λ	3.5	3.6	3.7	3.8	3.9	4.0	4.1	4.2	4.3	4.4
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	0.969803	0.972676	0.975276	0.977629	0.979758	0.981684	0.983427	0.985004	0.986431	0.987723
2	864112	874311	883799	892620	900815	908422	915479	922023	928087	933702
3	679153	697253	714567	731103	746875	761897	776186	789762	802645	814858
4	463367	484784	505847	526515	546753	566530	585818	604597	622846	640552
5	274555	293562	312781	332156	351635	371163	390692	410173	429562	448816
6	142386	155881	169912	184444	199442	214870	230688	246857	263338	280088
7	065288	073273	081809	090892	100517	110674	121352	132536	144210	156355
8	026739	030789	035241	040107	045402	051134	057312	063943	071032	078579
9	009874	011671	013703	015984	018533	021363	024492	027932	031933	037803
10	003315	004024	004848	005799	006890	008132	009540	011127	012906	014890
11	001019	001271	001572	001929	002349	002840	003410	004069	004825	005688
12	000289	000370	000470	000592	000739	000915	001123	001374	001566	002008
13	000076	000100	000130	000168	000216	000274	000345	000431	000534	000658
14	000019	000025	000034	000045	000059	000076	000098	000126	000160	000201
15	000004	000006	000008	000011	000015	000020	000026	000034	000045	000058
16	000001	000001	000002	000003	000004	000005	000007	000009	000012	000016
17				000001	000001	000001	000002	000002	000003	000004
18									000001	000001

表 3 正态分布

$$\Phi(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}e^{-\frac{x^2}{2}}dx\quad(u\geqslant 0)$$



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	u
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0.0
0.1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753	0.1
0.2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141	0.2
0.3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517	0.3
0.4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879	0.4
0.5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224	0.5
0.6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549	0.6
0.7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852	0.7
0.8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133	0.8
0.9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389	0.9
1.0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621	1.0
1.1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830	1.1
1.2	8849	8869	8888	8907	8925	8944	8962	8980	8997	90147	1.2
1.3	90320	90490	90658	90824	90988	91149	91309	91466	91621	91774	1.3
1.4	91924	92073	92220	92364	92507	92647	92785	92922	93056	93189	1.4
1.5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408	1.5
1.6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449	1.6
1.7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327	1.7
1.8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062	1.8
1.9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670	1.9

2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.0
2.1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574	2.1
2.2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899	2.2
2.3	98928	98956	98983	$9^{2}0097$	$9^{2}0358$	$9^{2}0613$	$9^{2}0863$	$9^{2}1106$	$9^{2}1344$	$9^{2}1576$	2.3
2.4	$9^{2}1802$	$9^{2}2024$	$9^{2}2240$	$9^{2}2451$	$9^{2}2656$	$9^{2}2857$	$9^{2}3053$	$9^{2}3244$	$9^{2}3431$	$9^{2}3613$	2.4
2.5	$9^{2}3790$	$9^{2}3963$	$9^{2}4132$	$9^{2}4297$	$9^{2}4457$	$9^{2}4614$	$9^{2}4766$	$9^{2}4915$	$9^{2}5060$	$9^{2}5201$	2.5
2.6	$9^{2}5339$	$9^{2}5473$	$9^{2}5604$	$9^{2}5731$	$9^{2}5855$	$9^{2}5975$	$9^{2}6093$	$9^{2}6207$	$9^{2}6319$	$9^{2}6427$	2.6
2.7	$9^{2}6533$	$9^{2}6636$	$9^{2}6736$	$9^{2}6833$	$9^{2}6928$	$9^{2}7020$	$9^{2}7110$	$9^{2}7197$	$9^{2}7282$	$9^{2}7365$	2.7
2.8	$9^{2}7445$	$9^{2}7523$	$9^{2}7599$	$9^{2}7673$	$9^{2}7744$	$9^{2}7814$	$9^{2}7882$	$9^{2}7948$	$9^{2}8012$	$9^{2}8074$	2.8
2.9	$9^{2}8134$	$9^{2}8193$	$9^{2}8250$	$9^{2}8305$	$9^{2}8359$	$9^{2}8411$	$9^{2}8462$	$9^{2}8511$	$9^{2}8559$	$9^{2}8605$	2.9
3.0	$9^{2}8650$	$9^{2}8694$	$9^{2}8736$	$9^{2}8777$	$9^{2}8817$	$9^{2}8856$	$9^{2}8893$	$9^{2}8930$	$9^{2}8965$	$9^{2}8999$	3.0
3.1	$9^{3}0324$	$9^{3}0646$	$9^{3}0957$	$9^{3}1260$	$9^{3}1553$	$9^{3}1836$	$9^{3}2112$	$9^{3}2378$	$9^{3}2636$	$9^{3}2886$	3.1
3.2	$9^{3}3129$	$9^{3}3363$	$9^{3}3590$	$9^{3}3810$	$9^{3}4024$	$9^{3}4230$	$9^{3}4429$	$9^{3}4623$	$9^{3}4810$	$9^{3}4991$	3.2
3.3	$9^{3}5166$	$9^{3}5335$	$9^{3}5499$	$9^{3}5658$	$9^{3}5811$	$9^{3}5959$	$9^{3}6103$	$9^{3}6242$	$9^{3}6376$	$9^{3}6505$	3.3
3.4	$9^{3}6631$	$9^{3}6752$	$9^{3}6869$	$9^{3}6982$	$9^{3}7091$	$9^{3}7197$	$9^{3}7299$	$9^{3}7398$	$9^{3}7493$	$9^{3}7585$	3.4
3.5	$9^{3}7674$	$9^{3}7759$	$9^{3}7842$	$9^{3}7922$	$9^{3}7999$	$9^{3}8074$	$9^{3}8146$	$9^{3}8215$	$9^{3}8282$	$9^{3}8347$	3.5
3.6	$9^{3}8409$	$9^{3}8469$	$9^{3}8527$	$9^{3}8583$	$9^{3}8637$	$9^{3}8689$	$9^{3}8739$	$9^{3}8787$	$9^{3}8834$	$9^{3}8879$	3.6
3.7	$9^{3}8922$	$9^{3}8964$	$9^{4}0039$	$9^{4}0426$	$9^{4}0799$	$9^{4}1158$	$9^{4}1504$	$9^{4}1838$	$9^{4}2159$	$9^{4}2468$	3.7
3.8	$9^{4}2765$	$9^{4}3052$	$9^{4}3327$	$9^{4}3593$	$9^{4}3848$	$9^{4}4094$	$9^{4}4331$	$9^{4}4558$	$9^{4}4777$	$9^{4}4988$	3.8
3.9	$9^{4}5190$	$9^{4}5385$	$9^{4}5573$	$9^{4}5753$	$9^{4}5926$	$9^{4}6092$	$9^{4}6253$	$9^{4}6406$	$9^{4}6554$	$9^{4}6696$	3.9
4.0	$9^{4}6833$	$9^{4}6964$	$9^{4}7090$	$9^{4}7211$	$9^{4}7327$	$9^{4}7439$	$9^{4}7546$	$9^{4}7649$	$9^{4}7748$	$9^{4}7843$	4.0
4.1	$9^{4}7934$	$9^{4}8022$	$9^{4}8106$	$9^{4}8186$	$9^{4}8263$	$9^{4}8338$	$9^{4}8409$	$9^{4}8477$	$9^{4}8542$	$9^{4}8605$	4.1
4.2	$9^{4}8665$	$9^{4}8723$	$9^{4}8778$	$9^{4}8832$	$9^{4}8882$	$9^{4}8931$	$9^{4}8978$	$9^{5}0226$	$9^{5}0655$	$9^{5}1066$	4.2
4.3	$9^{5}1460$	$9^{5}1837$	$9^{5}2199$	$9^{5}2545$	$9^{5}2876$	$9^{5}3193$	$9^{5}3497$	$9^{5}3788$	$9^{5}4066$	$9^{5}4332$	4.3
4.4	$9^{5}4587$	$9^{5}4831$	$9^{5}5065$	$9^{5}5288$	$9^{5}5502$	$9^{5}5706$	$9^{5}5902$	$9^{5}6089$	$9^{5}6268$	$9^{5}6439$	4.4
4.5	$9^{5}6602$	$9^{5}6759$	$9^{5}6908$	$9^{5}7051$	$9^{5}7187$	$9^{5}7318$	$9^{5}7442$	$9^{5}7561$	$9^{5}7675$	$9^{5}7784$	4.5
4.6	$9^{5}7888$	$9^{5}7987$	$9^{5}8081$	$9^{5}8172$	$9^{5}8258$	$9^{5}8340$	$9^{5}8419$	$9^{5}8494$	$9^{5}8566$	$9^{5}8634$	4.6
4.7	$9^{5}8699$	$9^{5}8761$	$9^{5}8821$	$9^{5}8877$	$9^{5}8931$	$9^{5}8983$	$9^{6}0320$	$9^{6}0789$	$9^{6}1235$	$9^{6}1661$	4.7
4.8	$9^{6}2067$	$9^{6}2453$	$9^{6}2822$	$9^{6}3173$	$9^{6}3508$	$9^{6}3827$	$9^{6}4131$	$9^{6}4420$	$9^{6}4696$	$9^{6}4958$	4.8
4.9	$9^{6}5208$	$9^{6}5446$	$9^{6}5673$	$9^{6}5889$	$9^{6}6094$	$9^{6}6289$	$9^{6}6475$	$9^{6}6652$	$9^{6}6821$	$9^{6}6981$	4.9

译名对照表

Banach	巴拿赫
Bayes	贝叶斯
Bernoulli	贝努里
Bertrand	贝特朗
Bochner	波赫纳
Boltzmann	波尔茨曼
Bolzano	波尔察诺
Borel	波雷尔
Boyle	波义耳
Brown	布朗
Buffon	浦丰
Cantelli	坎特里
Cauchy	哥西
Chapman	切普曼
Cramér	克拉美
De Mogan	隶莫根
De Moivre	隶莫佛尔
Euclid	欧几里德
Feller	费勒
Fourier	富里埃
Galton	高尔顿
Gamma	伽玛
Gauss	高斯
Hājek	哈依克
Helly	海莱
Herglotz	赫尔格洛茨
Jordan	约当
Laplace	拉普拉斯
Lebesgue	勒贝格
Lèvy	勒维
Lindeberg	林德伯格
Maxwell	麦克斯威尔
Pearson	皮尔逊
Pòlya	卜里耶
Poisson	泊松
Rayleigh	瑞利
Renyi	瑞尼
Riemann	黎曼
Schwarz	许瓦兹
Stieltjes	司蒂阶
Stirling	司蒂灵
Weibull	韦布
Weierstrass	维尔斯特拉斯
Колмогоров	柯尔莫哥洛夫
Ляпунов	李亚普诺夫
Марков	马尔科夫
Хинчин	辛钦
Чебышев	车贝谢夫

参考书目

[1] 概率论基础及其应用　王梓坤，科学出版社(1976)。

[2] 概率论与数理统计　复旦大学数学系，上海科学技术出版社(1961)。

[3] 概率论及数理统计　M. Fisz，王福保译，上海科学技术出版社(1962)。

[4] 概率论教程　Б. В. Гнеденко，丁寿田译，高等教育出版社(1956)。

[5] 统计数学方法　H. Cramer，魏宗舒等译，上海科学技术出版社(1966)。

[6] 相互独立随机变数之和的极限分布　Б. В. Гнеденко，А. Н. Голмогоров，王寿仁译，科学出版社(1955)。

[7] 微积分学教程　Р. М. Фихтенгольц　高等教育出版社(1954—1957)。

[8] 数学分析　江泽坚，上册：人民教育出版社(1964)，下册：高等教育出版社(1965)。

[9] 实变函数论　И. И. Натансон，徐瑞云译，上下册高等教育出版社(1955)。

[10] 复分析　L. V. Ahlfors　张立、张晴译，上海科学技术出版社(1984)。

[11] Probability, Random Variables and Stochastic Processes A. Papoulis, McGraw-Hill books Co. (1965).

[12] Introductory Probability and Statistical Applications P. L. Mayer, Oxford & IBH Pab. Co. (1970).
(有中译本：概率引论及统计应用，潘孝瑞等译，高等教育出版社(1986)。)

[13] Introductions to Probability and Statistics, PartI: Probability N. C. Giri Dekker (1974).

[14] Modern Probability Theory F. Z. Yeh, New York, Harper & Row. Pub (1973)

[15] 概率论导论及其应用　W. Feller，上册 胡迪鹤、林向清译，科学出版社(1964)，下册刘文译，科学出版社(1979)。

[16] 概率论　复旦大学信息论教研室(1979)
[17] 广义函数的基本理论

J. M. Mikusinski
R Sikorcki
复旦大学数学系译，人民教育出版社(1960)。

上册习题答案

第　一　章

1. (1) $\{2,3,4,5\}$,　　　　(2) $\{1,5,6,7,8,9,10\}$.

3. i) 提示：可利用互斥事件的想法(比如："装有 n 个球的罐中仅有一个红球，今从这罐中抽取 r 个球"这一事件，可分解为互斥事件之并)。

　ii) 利用 i)的结果。

4. i) 提示：设有一批产品共 n 件，其中 m 件为废品。现从中取 r 件，问这 r 件中有 x 件为废品的概率是多少？

　ii) (超几何分布)装有 n 个球的罐中有 m 个红球，今从罐中抽取 r 个球，求恰有 k 个红球的概率。

5. $\dfrac{2(n-r-1)}{n(n-1)}$.

6. 41/90.

7. $\dfrac{1}{2}$.

8. $\dfrac{\alpha}{\alpha+\beta}$.

9. i) $C_6^2 9^4/10^6$;　　　　ii) $A_{10}^6/10^6$;

　iii) $C_{10}^1C_6^2[A_9^4+C_9^1+C_9^1C_4^2C_8^1]/10^6$;　　　　iv) 为 ii)的逆事件。

10. $$P=1-\left[C_n^1\left(1-\frac{1}{n}\right)^k-C_n^2\left(1-\frac{2}{n}\right)^k+\cdots+\right.$$
$$\left.+C_n^{n-1}(-1)^n\left(1-\frac{n-1}{n}\right)^k\right].$$

11. i) $1-z$;　　　　ii) $y-z$;

　iii) $1-x+z$;　　　　iv) $1-x-y+z$.

12. 提示：注意 A、B 中至少有一个发生与 A、B 中仅有一个发生的关系。

13. $\dfrac{5}{8}$.

14. $\frac{1}{4}$.

15. 0.879.

17. 0.23.

18. 0.72.

20. 0.5103.

22. 提示：由一个装有 A 个球而其中有 a 个红球的罐子中任意取球而不放回去，试求早晚取到红球的概率。

23. 13/132.

24. 0.089.

25. $\frac{3}{4}$; $\frac{1}{2}$.

26. $\frac{25}{69}$; $\frac{28}{69}$; $\frac{16}{69}$.

27. 0.16.

29. $\frac{\alpha^2}{1-2\alpha\beta}$; $\frac{\beta^2}{1-2\alpha\beta}$.

31. $P(B)=0.595$.

32. 0.2

33. 1 次；32/81.

34. i) 0.8;　　ii) 0.252;　　iii) 0.85.

35. $p_1[1-(1-p_2p_3)(1-p)^2]$.

36. 提示：用 p_k 表示在 k 次独立试验中事件 A 出现偶数次的概率。利用互斥事件建立差分方程。$P(A)=\frac{1}{2}[1+(1-2p)^n]$.

38. e^{-at}.

第二章

6. $k=0.71$.

8. $$F(x)=\begin{cases}0, & x\leqslant 0;\\ \frac{x(2a-x)}{a^2}, & 0<x\leqslant a;\\ 1, & a<x.\end{cases}$$

12. 证 $\lim\limits_{x\to\infty} x\int_x^{\infty}\frac{1}{z}dF(z)\leqslant\lim\limits_{x\to\infty}\int_x^{\infty}dF(z)=0.$

类似可证

$$\lim_{x\to-\infty} x\int_{-\infty}^{x}\frac{1}{z}dF(z)=0.$$

现在证明

$$\lim_{x\to0+} x\int_{x}^{\infty}\frac{1}{z}dF(z)=0.$$

任给 $\alpha>\varepsilon>0$, 存在一个 $x_0>0$, 使

$$\int_{0+}^{x_0}dF(z)<\frac{\varepsilon}{2},$$

对于取定的 x_0, 只要 $0<x<\frac{x_0\varepsilon}{2}$, 就有

$$x\int_{x_0}^{\infty}\frac{1}{z}dF(z)\leqslant\frac{x}{x_0}\int_{x_0}^{\infty}dF(z)\leqslant\frac{x}{x_0}<\varepsilon/2,$$

$$x\int_{x}^{\infty}dF(z)=x\int_{x_0}^{\infty}\frac{1}{z}dF(z)+x\int_{x}^{x_0}\frac{1}{z}dF(z)$$

$$\leqslant\frac{x}{x_0}\int_{x_0}^{\infty}dF(z)+\int_{0+}^{x_0}dF(z)<\varepsilon/2+\varepsilon/2=\varepsilon.$$

故得证.

同理可证

$$\lim_{x\to0^-} x\int_{-\infty}^{x}\frac{1}{z}dF(z)=0.$$

13. 证 i) 显然, $\Phi(x)\geqslant0$.

ii) 因 $F(x)$ 是分布函数, 任给 $\varepsilon>0$, 存在 x_0, 使当 $x+h<x_0$ 时, 就有

$$F(x)<\varepsilon.$$

于是, 当 $x+h<x_0$ 时,

$$\Phi(x)=\frac{1}{h}\int_{x}^{x+h}F(z)dz<\frac{1}{h}\cdot\varepsilon h=\varepsilon,$$

即

$$\Phi(-\infty)=\lim_{x\to-\infty}\Phi(x)=0.$$

iii) 因 $F(x)$ 是分布函数, 任给 $\varepsilon_1>0$, 存在 x_1, 使当 $x_1<x$ 时, 就有

$$F(x)>1-\varepsilon_1.$$

于是，当 $x>x_1$ 时，

$$\Phi(x)=\frac{1}{h}\int_x^{x+h} F(z)dz>\frac{1}{h}(1-\varepsilon_1)\cdot h=1-\varepsilon_1$$

即

$$\Phi(+\infty)=\lim_{x\to+\infty}\Phi(x)=1.$$

iv) 证明 $\Phi(x)$ 单调不降.

任给 $\delta>0$，有

$$\Phi(x+\delta)=\frac{1}{h}\int_{x+\delta}^{x+h+\delta} F(z)dz=\frac{1}{h}\int_x^{x+h} F(z+\delta)dz$$

$$\geqslant\frac{1}{h}\int_x^{x+h} F(z)dz=\Phi(x).$$

即 $\Phi(x)$ 单调不降.

v) 证明 $\Phi(x)$ 左连续.

因为 $F(x)$ 左连续，任给 $\varepsilon>0$，存在 $\delta>0$，使当 $x-\delta\leqslant z<x$ 时，就有

$$F(x)-F(z)<\varepsilon$$

于是

$$\Phi(x)-\Phi(x-\delta)=\frac{1}{h}\int_x^{x+h} F(z)dz-\frac{1}{h}\int_{x-\delta}^{x-\delta+h} F(z)dz$$

$$=\frac{1}{h}\int_x^{x+h}[F(z)-F(z-\delta)]dz<\varepsilon.$$

由 $\Phi(x)$ 的单调不降性，当 $x-\delta\leqslant y<x$ 时，均有

$$\Phi(x)-\Phi(y)<\varepsilon,$$

即 $\Phi(x)$ 左连续.

由 i)--v) 可知，$\Phi(x)$ 为分布函数。

同理可证 $\Psi(x)$ 也是分布函数.

15. i) $g(x)=2x^2+\frac{2}{3}x,\quad 0\leqslant x\leqslant 1,$

$h(y)=\frac{y}{6}+\frac{1}{3},\qquad 0\leqslant y\leqslant 2;$

ii) $$\left.\begin{aligned} g(x|y)&=\frac{6x^2+2xy}{2+y} \\ h(y|x)&=\frac{3x+y}{6x+2}\end{aligned}\right\}\qquad\begin{pmatrix}0\leqslant x\leqslant 1\\0\leqslant y\leqslant 2\end{pmatrix}.$$

iii) 65/72; 7/24; 5/32

16. $f_{\xi_1}(x_1)=\frac{1}{\Gamma(k_1)}x_1^{k_1-1}e^{-x_1},\ 0<x_1<\infty$

$$f_{\xi_2}(x_2)=\frac{1}{\Gamma(k_1+k_2)}x_2^{k_1+k_2-1}e^{-x_2},\ 0<x_2<\infty.$$

18. i) ξ与η相互独立。

ii) ξ与η不独立。

19. $1-e^{-\frac{r^2}{2}}$.

20. $A=\frac{1}{2}$.

21. i) $F_\eta(y)=\begin{cases}0, & \text{当 } y\leqslant 0;\\ \frac{2\arcsin y}{\pi}, & \text{当 } 0<y\leqslant 1;\\ 1, & \text{当 } y>1.\end{cases}$

ii) $F_\eta(y)=\begin{cases}0, & \text{当 } y\leqslant 0;\\ 1-\frac{2}{\pi}\arccos y, & \text{当 } 0<y\leqslant 1;\\ 1, & \text{当 } y>1.\end{cases}$

23. ii) $P(\eta=j)=(2-q^j-q^{j+1})pq^j$;

iii) $P(\xi_1=k,\eta=j)=\begin{cases}pq^k(1-q^{k+1}), & \text{当 } k=j,\\ p^2q^{k+j}, & \text{当 } k<j,\\ 0, & \text{当 } k>j.\end{cases}$

25. i) $\xi+\eta$服从正态$N(0,\sqrt{2})$;

ii) $P(\xi-\eta=j)=\frac{1}{9.2^j}\sum_{k=0}^{j}C_2^kC_2^{j-k}2^k$;

iii) $f_{\xi+\eta}(z)=\begin{cases}0, & \text{当 } z\leqslant 0 \text{ 或 } z>3,\\ \frac{1}{2}z^2, & \text{当 } 0<z\leqslant 1,\\ -z^2+3z-\frac{3}{2}, & \text{当 } 1<z\leqslant 2,\\ \frac{1}{2}z^2-3z+\frac{9}{2}, & \text{当 } 2<z\leqslant 3;\end{cases}$

iv) $f_{\xi+\eta}(x)=\begin{cases}\dfrac{x+4}{24}, & \text{当} -4\leqslant x\leqslant 0,\\ \dfrac{1}{6}, & \text{当 } 0<x\leqslant 2,\\ \dfrac{6-x}{24}, & \text{当 } 2<x\leqslant 6,\\ 0, & \text{当 } x>6 \text{ 及 } x<-4;\end{cases}$

v) $f_{\xi+\eta}(x)=\dfrac{1}{2b}\left[\Phi\left(\dfrac{x-a+b}{\sigma}\right)-\Phi\left(\dfrac{x-a-b}{\sigma}\right)\right];$

vi) $f_{\xi+\eta}(z)=\begin{cases}0, & \text{当 } z<0,\\ e^{-\frac{z}{3}}\left(1-e^{-\frac{z}{6}}\right), & \text{当 } 0\leqslant z<\infty;\end{cases}$

vii) $F_{\xi+\eta}(x)=\dfrac{1}{2h}\displaystyle\int_{x-h}^{x+h}F(z)\,dz.$

26. $f_{\xi+\eta}(z)=ze^{-z},$

$f_{\frac{\xi}{\eta}}(z)=\dfrac{1}{(1+z)^2},$

$f_{\frac{\xi}{\xi+\eta}}(z)=\begin{cases}1, & 0<z<1;\\ 0, & \text{其它}.\end{cases}$

28. i) $f_{\xi\eta}(z)=\displaystyle\int_{-\infty}^{\infty}\frac{1}{|x|}f_\xi(x)f_\eta\left(\frac{z}{x}\right)dx,$

$f_{\xi-\eta}(z)=\displaystyle\int_{-\infty}^{\infty}f_\xi(x)f_\eta(x-z)\,dx;$

ii) $f_{\xi\eta}(z)=\begin{cases}\dfrac{1}{2a^2}\ln\dfrac{a^2}{|z|}, & \text{当 } |z|<a^2,\\ 0, & \text{当 } |z|\geqslant a^2,\end{cases}$

$f_{\xi-\eta}(z)=\begin{cases}0, & \text{当 } |z|>2a,\\ \dfrac{2a-z}{4a^2}, & \text{当 } 0\leqslant z\leqslant 2a,\\ \dfrac{2a+z}{4a^2}, & \text{当} -2a\leqslant z<0;\end{cases}$

iii) $f_{\xi\eta}(z)=\displaystyle\int_0^{\infty}\frac{1}{\pi x}e^{-\frac{x^4+z^2}{2x^2}}dx,$

$f_{\xi-\eta}(z)=\dfrac{1}{2\sqrt{\pi}}e^{-\frac{z^2}{4}}$. 即 $\xi-\eta$ 服从正态 $N(0,\ \sqrt{2})$。

29. $f_\zeta(z)=\begin{cases}-2\ln 4|z|, & \text{当} 0<|z|\leqslant\dfrac{1}{4};\\ 0, & \text{当} |z|>\dfrac{1}{4}.\end{cases}$

30. $\xi-\eta$ 服从正态 $N(0,\sqrt{\sigma_1^2+\sigma_2^2})$ 分布.

31. i) $f_{\frac{\xi}{\eta}}(z)=\dfrac{1}{\pi(1+z^2)}$;

ii) $f_{\frac{\xi}{\eta}}(z)=\begin{cases}0, & \text{当 } z\leqslant 0,\\ \dfrac{1}{2}, & \text{当 } 0<y\leqslant 1,\\ \dfrac{1}{2z^2}, & \text{当 } z>1.\end{cases}$

32. **证** 设 ξ 的密度函数为 $f(x)$，则 $(\xi,\ \eta)$ 的联合密度函数为 $f(x)\cdot f(y)$. 令

$$u=x+y,$$
$$v=x-y.$$

由定理 2.4.6，$(\xi+\eta,\xi-\eta)$ 的联合密度函数为

$$f\left(\frac{1}{2}(u+v),\frac{1}{2}(u-v)\right)\frac{1}{2}.$$

另一方面，由定理 2.4.1，$\xi+\eta$ 及 $\xi-\eta$ 的密度函数分别为

$$f_{\xi+\eta}(u)=\int_{-\infty}^{\infty}f(u-s)f(s)ds,$$

$$f_{\xi-\eta}(v)=\int_{-\infty}^{\infty}f(v+t)f(t)dt.$$

由 $\xi+\eta$ 与 $\xi-\eta$ 相互独立，得

$$\frac{1}{2}f\left(\frac{1}{2}(u+v)\right)f\left(\frac{1}{2}(u-v)\right)=f_{\xi+\eta}(u)f_{\xi-\eta}(v).$$

令 $g(u)=\lg f(u)$，于是

$$g\left(\frac{1}{2}(u+v)\right)+g\left(\frac{1}{2}(u-v)\right)=\lg_{\xi}f_{+\eta}(u)+\lg 2f_{\xi-\eta}(v).$$

因为 $g(u)$ 的二阶导数存在，上式对 u 求导得

$$g'\left(\frac{u+v}{2}\right)\frac{1}{2}+g'\left(\frac{u-v}{2}\right)\frac{1}{2}=(\lg f_{\xi+\eta}(u))'_u.$$

再对 v 求导，得

$$\frac{1}{4}g''\left(\frac{u+v}{2}\right)-\frac{1}{4}g''\left(\frac{u-v}{2}\right)=0.$$

对任意 x,y，选择 u,v 使

$$x=\frac{1}{2}(u+v),\quad y=\frac{1}{2}(u-v).$$

于是有

$$g''(x)-g''(y)=0.$$

由 x,y 的任意性得 $g''(x)\equiv$常数，因而

$$g(x)=a+bx+cx^2.$$

从而

$$f(x)=e^{a+bx+cx^2}.$$

故 ξ,η，从而 $\xi+\eta,\xi-\eta$ 皆服从正态分布。

33. 0.09.

34. 142克.

第 三 章

1. a) $E(\xi)=n;D(\xi)=n(n-1)$.

   b) $E(\xi)=\frac{n+1}{2};\ D(\xi)=\frac{n^2-1}{12}$

3. $E(\xi)=2;D(\xi)=2$.

4. $q=\frac{9}{10},\ P(\xi\leqslant 10)=1-\left(\frac{9}{10}\right)^{10}$.

5. $E(\xi)=8$.

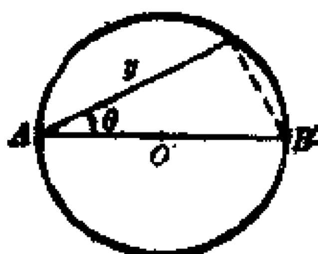
6. $E(\xi)=\frac{1}{b};\ D(\xi)=\frac{1}{b^2}$.

7. 提示：注意

$$\sum_{m=1}^{\infty}me^{-am}=-\frac{d}{da}\left(\sum_{m=0}^{\infty}e^{-am}\right);$$

$$\sum_{m=1}^{\infty}m^2e^{-am}=\frac{d^2}{da^2}\left(\sum_{m=0}^{\infty}e^{-am}\right);$$

$$E(\xi)=\frac{e^{-a}}{1-e^{-a}};D(\xi)=e^a/(e^a-1)^2.$$



8. $E(\xi)=0$; $D(\xi)=\frac{1}{2}$.

9. $E(\xi)=\frac{\alpha}{\alpha+\beta}$; $D(\xi)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

10. $E(\xi)=\sqrt{\frac{\pi a}{2}}$; $D(\xi)=\frac{(4-\pi)}{2}a$.

11. $E(\xi)=0$; $D(\xi)=2$.

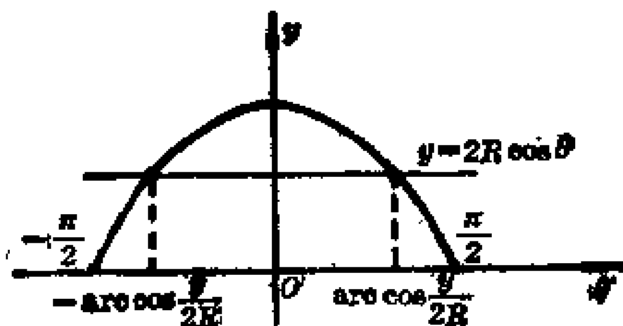
12. $E(\xi)=\frac{2a}{\sqrt{\pi}}$; $D(\xi)=\left(\frac{3}{2}-\frac{4}{\pi}\right)a^2$.

13. **解** 过已知点 A 作直径 $\overline{AB}$，则过 A 点的弦长 ξ 为：

$$\xi=2R\cos\theta,$$

θ 在 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 上均匀分布．于是，ξ 的分布函数为：

$$F_\xi(y)=\begin{cases}0, & y\leqslant 0;\\ P(2R\cos\theta<y), & 0\leqslant y\leqslant 2R;\\ 1, & y>2R.\end{cases}$$



第 13 题图

容易算得

$$P(2R\cos\theta<y)=\frac{2}{\pi}\arcsin\frac{y}{2R},$$

所以

$$f_\xi(y)=\begin{cases}\frac{2}{\pi}\frac{1}{\sqrt{4R^2-y^2}}, & 0<y<2R;\\ 0, & \text{其它},\end{cases}$$

$$E(\xi)=\frac{2}{\pi}\int_0^{2R}\frac{y}{\sqrt{4R^2-y^2}}dy=\frac{4R}{\pi}.$$

14. 提示：建立关于数学期望的微分方程式：

$$E[\xi(t)]=m_0e^{-pt}.$$

15. $E[\varphi]=\operatorname{arc\,tg}\frac{l}{h}+\frac{h}{l}\ln\left(\frac{h}{\sqrt{a^2+h^2}}\right).$

16. $E[\eta]=\frac{1}{2}[1-(q-p)^n].$

17. **解** 设 $P_n(A)$ 表示在 n 次试验后达到予定结果的概率，$P_{n,m}$ 为在 n 次试验中有 m 次成功的概率．由全概率公式，有

$$P_n(A)=\sum_{m=0}^{n}P_{n,m}G(m),$$

由题设条件

$$P_{n,m}=C_n^m p^m(1-p)^{n-m},$$

故

$$P_n(A)=\sum_{m=0}^{n}C_n^m p^m(1-p)^{n-m}\left\{1-\left(1-\frac{1}{\omega}\right)^m\right\}$$

$$=1-\left(1-\frac{p}{\omega}\right)^n.$$

如果在第 n 次试验时达到了予定的结果，则必须要有恰好 n 次的试验．此概率为 $P_n(A)-P_{n-1}(A)$．故为了达到予定结果所必须进行试验的次数的数学期望为

$$E(\xi)=\sum_{n=1}^{\infty}n[P_n(A)-P_{n-1}(A)]=\frac{\omega}{p}.$$

18. $E(\xi)=vp/q$; $D(\xi)=v\frac{p}{q}\left(1+\frac{p}{q}\right)$，其中 $q=1-p$.

23. $E[|\xi-a|]=\sqrt{\frac{2}{\pi}}\sigma.$

27. $P(\xi=k)=p^kq+pq^k$，$k=1,2,\cdots$;

$E(\xi)=\frac{p^2+q^2}{pq}$; $D(\xi)=\frac{q}{p^2}+\frac{p}{q^2}-2.$

28. $P(\eta=l)=p^2q^{l-1}+q^2p^{l-1}$，$l=1,2,\cdots$;

$E(\eta)=2$; $D(\eta)=2\left(\frac{q}{p}+\frac{p}{q}-1\right)$;

$P(\xi=k,\eta=l)=p^{k+1}q^l+q^{k+1}p^l.$

29. $E(\xi)=rp_1$，$D(\xi)=rp_1(1-p_1)$;

$E(\eta)=rp_2,\quad D(\eta)=rp_2(1-p_2)$;

$\operatorname{cov}(\xi,\eta)=-rp_1p_2$.

30. 证 (ξ,η)的联合密度函数是

$$f(x,y)=\frac{1}{2\pi\sqrt{1-R^2}}e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}$$

$$E[\max(\xi,\eta)]$$

$$=\frac{1}{2\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\max(x,y)\,e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}\,dxdy$$

$$=\frac{1}{2\pi\sqrt{1-R^2}}\left[\int_{-\infty}^{\infty}xdx\int_{-\infty}^{x}e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}\,dy\right.$$

$$\left.+\int_{-\infty}^{+\infty}ydy\int_{-\infty}^{y}e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}\,dx\right]$$

$$=\frac{1}{\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}xdx\int_{-\infty}^{x}e^{-\frac{1}{2(1-R^2)}[x^2-2Rxy+y^2]}\,dy$$

$$=\frac{1}{\pi\sqrt{1-R^2}}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{x}e^{-\frac{(y-Rx)^2}{2(1-R^2)}}dy$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}xe^{-\frac{x^2}{2}}dx\int_{-\infty}^{\sqrt{\frac{1-R}{1+R}}x}e^{-\frac{y^2}{2}}dy$$

$$=\frac{1}{\pi}\left[-e^{-\frac{x^2}{2}}\int_{-\infty}^{\sqrt{\frac{1-R}{1+R}}x}e^{-\frac{y^2}{2}}dy\right]_{-\infty}^{\infty}+\frac{1}{\pi}\sqrt{\frac{1-R}{1+R}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left[1+\frac{1-R}{1+R}\right]x^2}dx$$

$$=\sqrt{\frac{1-R}{\pi}}.$$

31. 提示:利用第二章习题 8 的结果。

$E(\xi)=l/3$; $D(\xi)=l^2/18$.

32. 提示:ξ与η的联合概率分布为

ξ \ η	0	1	2	3
0	0	0	$\frac{21}{120}$	$\frac{35}{120}$
1	0	$\frac{14}{120}$	$\frac{42}{120}$	0
2	$\frac{1}{120}$	$\frac{7}{120}$	0	0

$\rho=-0.58$.

33. 提示 利用第一章习题 21 的结果.

34. $E(\xi)=E(\eta)=\frac{\pi}{4}$;

$$D(\xi)=D(\eta)=\frac{\pi^2}{16}+\frac{\pi}{2}-2;$$

$$\operatorname{cov}(\xi,\eta)=\frac{\pi}{2}-1-\frac{\pi^2}{16}.$$

35. $\rho=\frac{n-m}{n}$.

37. 提示: 首先考虑下面的特殊情形:

$$\xi:\begin{pmatrix}0, & a\\ p, & q\end{pmatrix};\ \eta:\begin{pmatrix}0, & b\\ p', & q'\end{pmatrix}.$$

39. a) 证 设 $\max\limits_{1\le j\le k}x_j=x_{k_0}=a$, $x_j<a\ (j\ne k_0)$ $\quad 1\leqslant k_0\leqslant k$, 则

$$\frac{E(\xi^{n+1})}{E(\xi^n)}=\frac{\sum\limits_{j=1}^{k}x_j^{n+1}P(\xi=x_j)}{\sum\limits_{j=1}^{k}x_j^nP(\xi=x_j)}=\frac{a\sum\limits_{j=1}^{k}\left(\frac{x_j}{a}\right)^{n+1}P(\xi=x_j)}{\sum\limits_{j=1}^{k}\left(\frac{x_j}{a}\right)^{n}P(\xi=x_j)}$$

$$=\frac{a\left[P(\xi=x_{k_0})+\sum\limits_{j\ne k_0}\left(\frac{x_j}{a}\right)^{n+1}P(\xi=x_j)\right]}{P(\xi=x_{k_0})+\sum\limits_{j\ne k_0}\left(\frac{x_j}{a}\right)^{n}P(\xi=x_j)}\to a\ (\text{当 } n\to\infty).$$

类似地, 可以证明 b).

40. 证 $E(\xi)$存在, 即

$$\int_{-\infty}^{\infty}|x|\,dF(x)<\infty$$

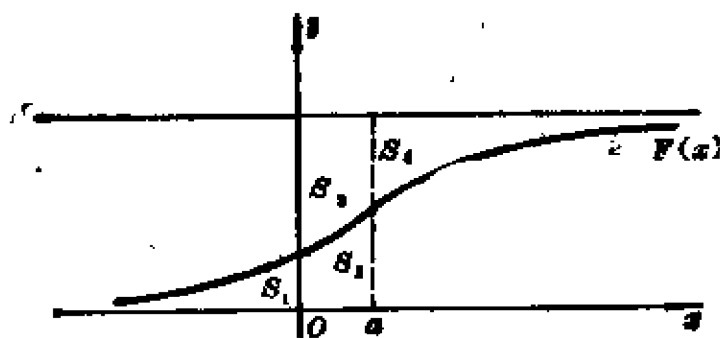
记 $E(\xi)=a$, 且设 $a>0$. 因为

$$\int_{-\infty}^{a}F(x)dx=xF(x)\Big|_{-\infty}^{a}-\int_{-\infty}^{a}xdF(x)$$

$$=aF(a)-\lim_{x\to-\infty}xF(x)-\int_{-\infty}^{a}xdF(x),$$

由于

$$\lim_{x\to-\infty}|xF(x)|=\lim_{x\to-\infty}\left|x\int_{-\infty}^{x}dF(y)\right|$$



第40题图

$$\leqslant \lim_{x\to-\infty}\left|\int_{-\infty}^{x} y\,dF(y)\right| \leqslant \lim_{x\to-\infty}\int_{-\infty}^{x}|y|\,dF(y)=0,$$

故

$$\int_{-\infty}^{a}F(x)\,dx=aF(a)-\int_{-\infty}^{a}x\,dF(x).$$

同理可证

$$\int_{a}^{\infty}[1-F(x)]\,dx=-\int_{-\infty}^{a}x\,dF(x)+aF(a).$$

故

$$\int_{-\infty}^{a}F(x)\,dx=\int_{a}^{\infty}[1-F(x)]\,dx.\quad (*)$$

记

$$S_1=\int_{-\infty}^{0}F(x)\,dx,\qquad S_2=\int_{0}^{a}F(x)\,dx,$$

$$S_3=\int_{0}^{a}[1-F(x)]\,dx,\qquad S_4=\int_{a}^{\infty}[1-F(x)]\,dx.$$

(*)式说明 $S_1+S_2=S_4$(见上图). 故

$$\begin{aligned}&\int_{0}^{\infty}[1-F(x)]\,dx-\int_{-\infty}^{0}F(x)\,dx=S_3+S_4-S_1\\&=S_3+S_4-S_4+S_2=S_2+S_3\\&=\int_{0}^{a}F(x)\,dx+\int_{0}^{a}[1-F(x)]\,dx\\&=\int_{0}^{a}dx=a.\end{aligned}$$

对于 $a<0$ 的情形,可类似证明.

41. $E\left[\xi\mid\eta=\frac{1}{2}\right]=\frac{7}{9}$; $D\left[\xi\mid\eta=\frac{1}{2}\right]=13/648$.

42. 解　依题意，当 $N=n$ 时，ξ 服从参数为 $n, p=0.10$ 的二项分布，因而 $E[\xi|N]=0.10N$. 由全期望公式

$$E(\xi)=E[E(\xi|N)]=E[0.10N]=0.10E(N).$$

但

$$E(N)=13.3$$

故

$$E(\xi)=0.10\times 13.3=1.33.$$

第　四　章

1. $\varphi(t)=\dfrac{p}{1-qe^{jt}}$; $E(\xi)=q/p$;

$$D(\xi)=\frac{q}{p}+\frac{q^2}{p^2}.$$

3. ξ: $\begin{pmatrix} \cdots, & -n, & \cdots, & -1, & 0, & 1, & \cdots, & n, & \cdots \\ \cdots, & \dfrac{a_n}{2}, & \cdots, & \dfrac{a_1}{2}, & a_0, & \dfrac{a_1}{2}, & \cdots, & \dfrac{a_n}{2}, & \cdots \end{pmatrix}$,

η: $\begin{pmatrix} 0, & 1, & 2, & \cdots, & n, & \cdots \\ a_0, & a_1, & a_2, & \cdots, & a_n, & \cdots \end{pmatrix}$

5. 提示：$F(\xi)$ 为$[0,1]$上的均匀分布。

a) $\varphi(t)=\dfrac{e^{jat}-1}{jat}e^{jbt}$;

b) $\varphi(t)=\dfrac{1}{1+jt}$.

6. 提示：由第二章习题 13，$G(x)$ 也是分布函数，并且 $G(x)$ 是 $F(x)$ 与 $[-h,h]$ 上均匀分布的卷积.

9. 提示：实值特征函数可写为：

$$\varphi(t)=\int_{-\infty}^{\infty}\cos tx\,dF(x).$$

10. 提示：证明 $\varphi(t)$ 是离散型随机变数

$$P\left(\xi=\frac{n\pi}{a}\right)=p_n,\quad n=0,\ \pm1,\cdots$$

的特征函数. 其中 $p_0=\dfrac{1}{2}$, $p_{-(2k-1)}=p_{2k-1}=\dfrac{2}{\pi^2}\dfrac{1}{(2k-1)^2}$, $k=1,2,\cdots$.

13. $C=\dfrac{\sqrt{76}}{2\pi}$, $E(\xi)=5$, $E(\eta)=3$.

$$\varphi(t_1, t_2)=e^{i(5t_1+3t_2)}e^{-\left[\frac{5}{76}t_1^2-\frac{3}{76}t_1t_2+\frac{1}{19}t_2^2\right]}.$$

14. a) $\psi_\xi(s)/(1-s)$;

b) $s\psi_\xi(s)/(1-s)$;

c) $(1-s\psi_\xi(s))/(1-s)$;

d) $\dfrac{p_0}{s}+\dfrac{s-\psi_\xi(s)}{s(1-s)}$;

e) $\dfrac{1}{2}[\psi_\xi(s^{\frac{1}{2}})+\psi_\xi(-s^{\frac{1}{2}})]$.

15. 提示：建立关于 u_n 的差分方程。

$$\psi(s)=\frac{pqs^2}{(1-ps)(1-qs)};$$

$$E(\xi)=\frac{1}{pq},\ D(\xi)=\frac{1-3pq}{p^2q^2}.$$

16. $\psi_{\tau_r}(s)=\left(\dfrac{q}{1-ps}\right)^r$.

17. $E(R)=1+\dfrac{q\nu}{p}$; $D(R)=\dfrac{q\nu}{p^2}$.

19. $\psi(s)=\dfrac{1}{1-qs}+\dfrac{p^2s^2}{(1-qs)^2}\psi(s)$.

第五章

1. 提示：利用$\dfrac{1}{x}e^{-\frac{x^2}{2}}=\displaystyle\int_x^\infty\left(1+\frac{1}{y^2}\right)e^{-\frac{y^2}{2}}dy$.

5. 1) 提示：因为

$$|\xi_n^2-a^2|\leqslant|\xi_n-a|^2+2|a||\xi_n-a|.$$

所以，当 $a\neq0$ 时，对任给 $\varepsilon>0$，有

$$P(|\xi_n^2-a^2|\geqslant2\varepsilon)\leqslant P(|\xi_n-a|^2\geqslant\varepsilon)+P\left(|\xi_n-a|\geqslant\frac{\varepsilon}{2|a|}\right).$$

2) 提示：由于$\lim\limits_{n\to\infty}P\left(|\eta_n|<\left|\dfrac{b}{2}\right|\right)=0$，故

$$\lim_{n\to\infty}P\left\{\left|\frac{\xi_n}{\eta_n}-\frac{a}{b}\right|\geqslant\varepsilon\right\}$$

$$\leqslant\lim_{n\to\infty}P\left\{|b(\xi_n-a)-a(\eta_n-b)|\geqslant\frac{b^2}{2}\varepsilon\right\}+\lim_{n\to\infty}P\left(|\eta_n|\leqslant\frac{|b|}{2}\right).$$

7. 提示：注意，任给 $\varepsilon>0$，

$$E[f(|\xi_n|)]=\int\limits_{|x|>\varepsilon} f(|x|)\,dF_n(x)+\int\limits_{|x|\leqslant\varepsilon} f(|x|)\,dF_n(x).$$

9. 证 设 $F_n(x)\ (n=1,2,\cdots)$ 及 $F(x)$ 分别为 ξ_n 及 ξ 的分布函数，并设 x_0 为 $F(x)$ 的连续点. 则

$$\lim_{n\to\infty}F_n(x_0)=F(x_0),$$

且任给 $\varepsilon>0$，存在 $\delta>0$，使当 $|x-x_0|<\delta$ 时，

$$|F(x)-F(x_0)|<\varepsilon.$$

因为 $F(x)$ 在 $(x_0-\delta,\ x_0+\delta)$ 内最多只有可列个间断点. 故存在 $F(x)$ 的连续点 x_1 及 x_2 满足

$$x_0-\delta<x_1<x_0<x_2<x_0+\delta$$

令 $\delta^*=\min(x_0-x_1, x_2-x_0)$，由题设条件，存在 N，使当 $n>N$ 时，

$$|F_n(x_1)-F(x_1)|<\varepsilon,$$
$$|F_n(x_2)-F(x_2)|<\varepsilon,$$
$$P(|\eta_n|\geqslant\delta^*)<\varepsilon.$$

因而

$$\{\xi_n<x_0-\delta^*\}-\{\xi_n<x_0-\delta^*,\ |\eta_n|\geqslant\delta^*\}$$
$$=\{\xi_n<x_0-\delta^*,\ |\eta_n|<\delta^*\}\subset\{\xi_n<x_0+\delta^*\}$$

所以，当 $n\geqslant N$ 时，

$$|P(\xi_n+\eta_n<x_0,\ |\eta_n|<\delta^*)-F(x_0)|<3\varepsilon,$$

从而

$$|P(\xi_n+\eta_n<x_0)-F(x_0)|<4\varepsilon.$$

即

$$\xi_n+\eta_n\xrightarrow{W}\xi.\ ■$$

11. 提示：令

$$\xi_k=\begin{cases}1, & \text{第 } k \text{ 号球投入第 } k \text{ 个匣子;}\\ 0, & \text{其它,}\end{cases}$$

$$S_n=\sum_{k=1}^{n}\xi_k.$$

12. 提示：对任给 $\delta>0$，取 $M>0$ 及正整数 N_1，使当 $n>N_1$ 时有

$$P(|\xi|\geqslant M)<\frac{1}{3}\delta,$$

$$P(|\xi_n-\xi|\geqslant M)<\frac{1}{3}\delta,$$

由于 $g(x)$在$[-2M,2M]$上一致连续，故对任给 $\varepsilon>0$，存在 $\varepsilon_1>0$，使当 x_1，$x_2\in[-2M,2M]$且 $|x_1-x_2|<\varepsilon_1$ 时有

$$|g(x_1)-g(x_2)|<\varepsilon.$$

取正整数 $N\geqslant N_1$，使当 $n>N$ 时有

$$P(|\xi_n-\xi|\geqslant\varepsilon_1)<\frac{1}{3}\delta.$$

注意到

$$\{|\xi|\geqslant 2M\}\cup\{|\xi_n|\geqslant 2M\}\subset\{|\xi_n-\xi|\geqslant\varepsilon_1\},$$

所以当 $n>N$ 时，有

$$\begin{aligned}&P(|g(\xi_n)-g(\xi)|\geqslant\varepsilon)\leqslant P(|\xi|\geqslant 2M)\\&+P(|\xi_n|\geqslant 2M)+P(|\xi|<2M,\ |\xi_n|<2M,|g(\xi_n)-g(\xi)|\geqslant\varepsilon)\\&\leqslant\delta.\end{aligned}$$

19. 辛钦大数定律不成立。

21. a) 大数定律不成立.

 b) 强大数定律成立.

 c) 大数定律不成立.

25. 提示：采用证明辛钦大数定律时所用的方法. 任给 $\delta>0$，对于 $k=1,2,\cdots$，令

$$\eta_{nk}=\xi_k,\zeta_{nk}=0,\quad 当|\xi_k|<\delta n;$$
$$\eta_{nk}=0,\zeta_{nk}=\xi_k,\quad 当|\xi_k|\geqslant\delta n,$$

则

$$\xi_k=\eta_{nk}+\zeta_{nk}.$$

28. 提示：利用概率连续性定理。

33. 强大数定律成立.

36. b) 不服从中心极限定理.

37. 满足中心极限定理.

38. 满足中心极限定理.

40. 中心极限定理成立的条件是：

 i) a_n 一致有界且 $S_n\to\infty$；

或

ii) $\sum_{k=n}^{\infty} a_k^2$

$\sum_{k=1} p_k a_k^2$.

大数定律成立的条件是：

i) a_n 一致有界，

或

ii) 存在 $\delta>0$，使 a_k^2 与 $k^{1-\delta}$ 为同级无穷大。

41. 李亚普诺夫定理成立.

43. 提示：利用第 36 题 a)的结果去证明必要性。

目　录

第六章 抽 样 分 布

前述各章，是概率论的引论，讲述了概率论的基本内容，为数理统计学建立了重要的数学基础. 从本章起的接连四章，是数理统计学的初步，主要讲述估计与检验等原理，回归分析与方差分析等统计方法.

数理统计学是运用概率论的基本知识，对要研究的随机现象进行多次观察或试验，研究如何合理地获得数据资料，建立有效的数学方法，根据所获得的数据资料，对所关心的问题作出估计与检验. 数理统计学的重要分支有统计推断、多元统计分析、试验设计等，其具体方法甚多，应用相当广泛，已成为各学科从事科学研究及生产、经济等部门进行有效工作的必不可少的数学工具.

在这一章中，我们从数理统计学的基本概念讲起，讨论抽样分布及其重要的定理. 这些抽样分布及其几个重要的定理，在前述各章中尚未提到，而在后述三章中却经常要用到它们.

§6.1 基 本 概 念

一、总体、个体、简单随机子样

总体、个体、子样是数理统计学中三个最基本的术语. 我们把对某一个问题的研究对象的全体称为总体(或母体)，组成总体的每个基本单元称为个体，从总体中随机抽取的 n 个个体称为容量为 n 的子样. 例如，把某月的整批产品视为一总体，则每个产品为个体；又如把某批灯泡视为一总体，则每个灯泡为个体；某地在某季度内每天的日平均气温的全体视为一总体，则其中某天的日平

均气温为个体 在数理统计学中，我们是对总体成员的一个或者若干个数量表征进行研究. 如日平均气温的度数用ξ表示，灯泡的使用寿命(小时)用η表示，这样对总体的研究就归结为讨论随机变数ξ或η的分布函数及其主要数字特征(如数学期望$E(\xi)$、方差$D(\xi)$等)的研究. 又如某月的整批产品为一总体，若只要检查产品的质量，可用数量表征ξ来反映. 产品为一等品，令ξ取值为1；二等品，ξ取值为2；不合格品，ξ取值为3. 我们研究这批产品的质量规律，就归结为讨论随机变数ξ的分布函数及其主要数字特征了

今后我们常用"总体ξ服从什么分布"这样的术语，它是指总体的某个具体数量表征ξ服从什么分布规律. 我们说对总体进行n次独立的重复试验或n次独立的观察，就是从总体中随机地抽取容量为n的子样，对应的数量表征以$\xi_1, \cdots, \xi_n$表示，不难理解，ξ及$\xi_1, \cdots, \xi_n$都是随机变数，$(\xi_1, \cdots, \xi_n)$是n维随机向量. 例如我们说整批灯泡的寿命分布如何，是指灯泡的寿命ξ服从什么分布. 从整批灯泡中随机抽取容量为n的子样，是指随机抽取n个灯泡，观察这n个灯泡的寿命，用$\xi_1, \cdots, \xi_n$表示，其中ξ_i表示随机抽取到的第i个灯泡的寿命，显然ξ_i也是一随机变数$(i=1, \cdots, n)$. 为区别总体与子样的概念，我们今后常用诸如"总体ξ服从什么分布"、"从总体ξ中抽取子样$\xi_1, \cdots, \xi_n$"等这类术语.

我们在对总体进行研究中，当总体相当大时（指个体相当多时)或当对个体进行试验具有破坏性(如炮弹能否引发爆炸) 或费时耗资大时，只可能从总体中抽取部分个体来作研究，即抽取容量为n的子样来作研究. 我们要从子样的观察或试验结果的特性来对总体的特性作出估计与推断，一方面自然要研究应该怎样从总体中抽取子样，使得子样在尽可能大的程度上反映总体的特性；同时必须建立一整套的方法，使能根据所选取的子样的性质，来对总

体的特性进行估计与推断．因此，我们在抽取子样来对总体作出估计与推断时，从总体中抽取子样必须是随机的，即每一个体都有同等概率被抽取（当总体中的个体是有限个时，要用有返回抽取方式）．其具体要求为两个方面：独立性，是指$\xi_1,\cdots,\xi_n$相互独立；代表性，是指$\xi_1,\cdots,\xi_n$中每一个都与总体ξ有相同分布．

定义 6.1.1 （简单随机子样） 设$\xi_1,\cdots,\xi_n$为来自总体ξ的容量为n的子样，如果$\xi_1,\cdots,\xi_n$相互独立且每一个都是与总体ξ有相同分布的随机变数，则称$\xi_1,\cdots,\xi_n$为总体ξ的容量为n的简单随机子样，简称为简单子样或子样①．

在这一章中，我们所讲的子样$\xi_1,\cdots,\xi_n$，无特别声明的话，都是指简单子样，即$\xi_1,\cdots,\xi_n$相互独立且每一个都与总体ξ有相同的分布．引入简单随机子样，是基于要从子样所获得的概率统计特性来对总体的概率统计特性作出估计与推断．对于简单随机子样，我们可以应用概率论中对独立随机变数的情形所建立的许多重要的定理，这些重要的结论为数理统计学提供了必要的基础．

二、统计量

子样$\xi_1,\cdots,\xi_n$也可用n维随机向量$(\xi_1,\cdots,\xi_n)$表示．记x_i为ξ_i的一次观察值，并称$(x_1,\cdots,x_n)$为子样的一次观察值．子样$(\xi_1,\cdots,\xi_n)$的所有可能取值的全体称为子样空间，记作$\mathscr{X}$，它是n维空间．子样的一次观察值$(x_1,\cdots,x_n)$就是子样空间$\mathscr{X}$中的一个点，即$(x_1,\cdots,x_n)\in\mathscr{X}$．

定义 6.1.2（统计量） 设$\xi_1,\cdots,\xi_n$为总体ξ的子样，T为子样空间$\mathscr{X}$中点$(x_1,\cdots,x_n)$的实值函数，作子样的函数$T=T(\xi_1,$

① 是指若ξ为定义在概率空间$(\Omega,\mathscr{F},P)$上的随机变数，则$\xi_1,\cdots,\xi_n$为概率空间$(\Omega,\mathscr{F},P)$上n个相互独立的并同ξ具有相同分布的随机变数．

$\cdots,\xi_n)$，T 的取值记为 $t=T(x_1,\cdots,x_n)$．若 $T(\xi_1,\cdots,\xi_n)$ 也为一随机变数①，且不带未知参数，则称 T 或 $T(\xi_1,\cdots,\xi_n)$ 为统计量．

我们在用子样 $\xi_1,\cdots,\xi_n$ 获得的信息来对总体 ξ 作出估计与推断时，是按不同的统计问题的要求而规定子样的各种函数．在这本教材中，所涉及到的子样的各种函数 $T(\xi_1,\cdots,\xi_n)$，$T(x_1,\cdots,x_n)$ 一般都是多维随机变量的连续函数，因而 $T(\xi_1,\cdots,\xi_n)$ 都是随机变数．

例 6.1.1 设 $\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，其容量为 n．记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\qquad S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2,$$

则 $\bar{\xi}$ 及 S^2 都是统计量，称 $\bar{\xi}$ 及 S^2 分别为子样 $\xi_1,\cdots,\xi_n$ 的平均值及方差．子样的观察值为 $x_1,\cdots,x_n$，$\bar{\xi}$ 及 S^2 的观察值分别记作

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,\qquad s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

今后，大写的 S^2 表示统计量，小写的 s^2 表示统计量 S^2 的观察值．

定义 6.1.3（顺序统计量） 设 $\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，今由子样建立 n 个函数：

$$\xi_k^*=\xi_k^*(\xi_1,\cdots,\xi_n)\qquad k=1,\cdots,n,$$

其中 ξ_k^* 为这样的统计量，它的观察值为 x_k^*，x_k^* 为子样 $\xi_1,\cdots,\xi_n$ 的观察值 $x_1,\cdots,x_n$ 中由小至大排列（即 $x_1^*\leqslant\cdots\leqslant x_k^*\leqslant\cdots\leqslant x_n^*$）后的第 k 个数值，则称 $\xi_1^*,\cdots,\xi_n^*$ 为顺序统计量．

易见，$\xi_1^*=\min\{\xi_1,\cdots,\xi_n\}$，$\xi_n^*=\max\{\xi_1,\cdots,\xi_n\}$．称 ξ_1^* 为最小项统计量，ξ_n^* 为最大项统计量．若 n 为奇数，则称 $\xi^*_{\frac{n+1}{2}}$ 为子样

---

① 具体要求就是 $T(\xi_1,\cdots,\xi_n)$ 是 n 元可测实函数，也就是在第二章中所述的 n 元随机变数，它满足对每一 $t\in R_1$，有 $\{T(\xi_1,\cdots,\xi_n)\leqslant t\}\in\mathscr{F}$．

的中值；若 n 为偶数，则称 $\xi^*_{\frac{n}{2}+1}$ 为子样的中值.

定义 6.1.4(极差) 设 $\xi_1, \cdots, \xi_n$ 为总体 ξ 的子样，则称统计量 $D^*_n \triangleq \xi^*_n - \xi^*_1$ 为子样的极差.

它是子样中最大值与最小值之差，反映了子样观察值的波动幅度. 它同方差一样是反映观察值离散程度的数量指标，而且计算方便.

例 6.1.2 设 $\xi_1, \cdots, \xi_5$ 为 ξ 的容量为 5 的子样，今对这个子样作了三次观察，其值如表 6.1.1 所示，试求 $\bar{\xi}$、S^2 及 D^*_n 的观察值.

表 6.1.1

ξ	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5
1	3	1	10	5	6
2	2	6	7	2	8
3	8	3	9	10	5

表 6.1.2

ξ^*_1	ξ^*_2	ξ^*_3	ξ^*_4	ξ^*_5	$\bar{x}$	s^2	D^*_n
1	3	5	6	10	5	9.2	9
2	2	6	7	8	5	6.4	6
3	5	8	9	10	7	6.8	7

三、小样问题与大样问题

统计量是我们对总体 ξ 的分布函数或数字特征进行估计与推断最重要的基本概念，求出统计量 $T(\xi_1, \cdots, \xi_n)$ 的分布函数是数理统计学的基本问题之一. 统计量的分布，称为抽样分布.

设总体 ξ 的分布函数表达式已知，对于任一自然数 n，如能求

出给定统计量 $T(\xi_1,\cdots,\xi_n)$ 的分布函数，这分布称为统计量 T 的精确分布. 求出统计量 T 的精确分布，这对于数理统计学中的所谓小样问题（即在子样容量 n 比较小的情况下所讨论的各种统计问题）的研究是很重要的.

但一般说来，要确定一个统计量的精确分布其难度比较大.只对一些重要的特殊情形，如总体 ξ 服从正态分布时，已求出 t 统计量、χ^2 统计量、F 统计量等的精确分布. 它们在参数的估计及检验中起很重要的作用.

若统计量 $T(\xi_1,\cdots,\xi_n)$ 的精确分布求不出来，或其表达式非常复杂而难于应用，但如能求出它在 $n\to\infty$ 时的极限分布，那么这个统计量的极限分布对于数理统计学中的所谓大样问题（即在子样容量 n 比较大的情况下讨论的各种统计问题）的研究很有用. 但要注意，在应用极限分布时，要求子样的容量 n 比较大. 如第八章的 § 8.3 所讨论的非参数性假设检验问题，是用检验统计量的极限分布，因而子样容量 n 应取得比较大才行.

§ 6.2　子样的数字特征及其分布

*一、经验分布与格列汶科定理

在实际工作中遇到各种各样的随机变数，怎样确定它的分布函数 $F(x)$？在概率论中，我们介绍了常用的几种分布函数以及它们的一些性质，在那里我们假定它们都是事先给定了的. 现在，我们介绍用独立重复试验的方法，利用子样建立一定的概率模型，用由此所获得的概率统计特性来对总体 ξ 的分布函数 $F(x)$ 等作出估计与推断.

定义6.2.1（经验分布函数）　从总体 ξ 中抽取容量为 n 的子样 $\xi_1,\cdots,\xi_n$，当顺序统计量 $\xi_1^*,\xi_2^*,\cdots,\xi_n^*$ 的值给定时，对任何实数 x，我们定义函数 $F_n^*(x)$：

$$F_n^*(x)=\begin{cases}0, x\leqslant x_1^*\\ \frac{k}{n}, x_k^*<x\leqslant x_{k+1}^*, k=1,\cdots,n-1,\\ 1, x>x_n^*\end{cases} \qquad (6.2.1)$$

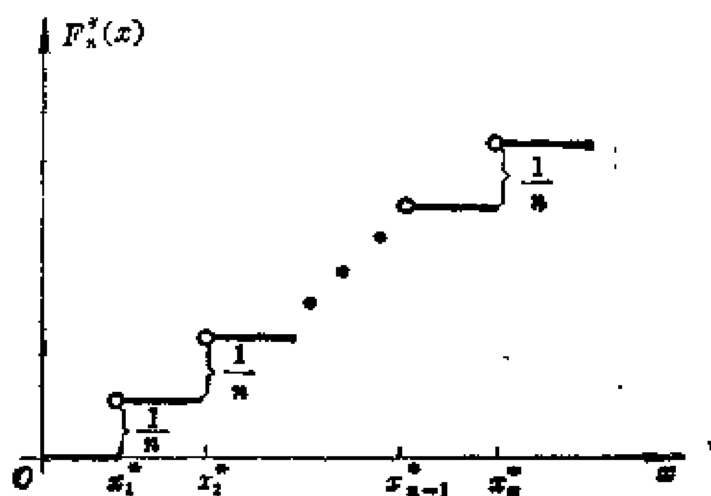


图 6.2.1

称 $F_n^*(x)$ 为总体 ξ 的经验分布函数①.

易见，对于每一组观察值 $\xi_i^*=x_i^*, i=1,\cdots,n, F_n^*(x)$ 单调、非降、左连续且在 $x=x_k^*$ 点有间断点，在每个间断点上跳跃量都是 $\frac{1}{n}$. 显然，$0\leqslant F_n^*(x)\leqslant 1$，并具有分布函数的其它性质.

由定义 6.2.1 知，对于 x 的每一数值而言，经验分布函数 $F_n^*(x)$ 为子样 $\xi_1,\cdots,\xi_n$ 的函数，它是一统计量，即为一随机变数，其可能取值为 $0,\frac{1}{n},\cdots,\frac{n-1}{n}$, 1. 事件"$F_n^*(x)=\frac{k}{n}$"发生的概率，由于 $\xi_1,\cdots,\xi_n$ 相互独立且有相同的分布函数 $F(x)$，因而它等价于

① 对于每一 ω，即子样的一次观察值，由(6.2.1)定义的 $F_n^*(x)$ 是实数 x 的函数，$x\in R_1$. 而对于每一 x 值，$F_n^*(x)$ 依赖于子样的观察值，即为 ω 的函数，因而 $F_n^*(x)$ 为一统计量. 这里，如果 $x_k<x_{k+1}=x_{k+2}=\cdots$，而 $x_k<x=x_{k+1}=x_{k+2}=\cdots$，则应在(6.2.1)中把 $\frac{k}{n}$ 改为 $\frac{N_k(x)}{n}$，其中 $N_k(x)$ 是子样中小于 x_{k+1} 的观察值的个数.

n 次独立重复试验的贝努里概型中事件"$\xi<x$"发生 k 次而其余 $n-k$ 次不发生的概率，即有：

$$P\left\{F_n^*(x)=\frac{k}{n}\right\}=C_n^k\{F(x)\}^k\{1-F(x)\}^{n-k} \qquad (6.2.2)$$

其中 $F(x)=P(\xi<x)$，它是总体 ξ 的分布函数.

我们从第五章所讲的大数定律知道，在一定的条件下，事件发生的频率依概率收敛于这个事件发生的概率. 人们自然要问，总体 ξ 的经验分布函数 $F_n^*(x)$，即"$\xi<x$"事件发生的频率，当 n 足够大时，是否也渐近于事件"$\xi<x$"发生的概率，即总体 ξ 的分布函数 $F(x)$ 呢？格列汶科于 1933 年作了肯定的回答.

格列汶科定理 设总体 ξ 的分布函数为 $F(x)$，经验分布函数为 $F_n^*(x)$，对于任何实数 x，记

$$D_n=\sup_{-\infty<x<+\infty}|F_n^*(x)-F(x)|, \qquad (6.2.3)$$

则有

$$P\{\lim_{n\to\infty}D_n=0\}=1. \qquad (6.2.4)$$

我们知道 $F_n^*(x)$ 为一统计量，因而 D_n 也为一统计量. D_n 这个统计量用来衡量 $F_n^*(x)$ 同 $F(x)$ 之间在所有的 x 值上最大的差异程度. 格列汶科定理证明了统计量 D_n 以概率为 1 地收敛于零. 通俗地说，就是当 n 足够大时，对于所有的 x 值，$F_n^*(x)$ 同 $F(x)$ 之差的绝对值都很小这个事件发生的概率等于 1. 但是格列汶科定理还未阐明统计量 D_n 服从什么分布或以什么分布为其极限分布. 我们将在第八章的 §8.3 中叙述两个重要定理，即柯尔莫哥洛夫定理和斯米尔诺夫定理. 这两个定理找到了统计量 D_n 的极限分布表达式，这些定理在分布函数的假设检验中有重要的作用.

二、子样的数字特征

1. 子样矩

设 $F_n^*(x)$ 为总体 ξ 的经验分布函数，记

$$A_r=\int_{-\infty}^{\infty}x^r dF_n^*(x)=\frac{1}{n}\sum_{i=1}^{n}\xi_i^r, \tag{6.2.5}$$

$$B_r=\int_{-\infty}^{\infty}(x-\bar{\xi})^r dF_n^*(x)=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^r. \tag{6.2.6}$$

称 A_r 及 B_r 分别为子样的 r 阶原点矩及 r 阶中心矩，其中 r 为正整数. 显然 $A_1=\bar{\xi}$，$B_2=S^2$. 由于 A_r 及 B_r 都是子样的连续函数，因而都是统计量，也即都是随机变数.

2. 协方差与相关系数

设 $(\xi_1,\eta_1),\cdots,(\xi_n,\eta_n)$ 为二维总体 (ξ,η) 的子样，记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i \qquad S_1^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$$

$$\bar{\eta}=\frac{1}{n}\sum_{i=1}^{n}\eta_i \qquad S_2^2=\frac{1}{n}\sum_{i=1}^{n}(\eta_i-\bar{\eta})^2 \tag{6.2.7}$$

$$S_{12}=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\eta_i-\bar{\eta})$$

称统计量 S_{12} 为子样的协方差，称统计量

$$R=\frac{S_{12}}{S_1\cdot S_2} \tag{6.2.8}$$

为子样的相关系数.

三、子样数字特征的分布

我们在这一段叙述子样的平均值、极值与极差的分布，子样相关系数 R 的分布将在§8.3 中叙述，子样方差 S^2 的分布在下一节中叙述.

1. 子样平均值的分布

设 $\varphi(t)$ 为总体 ξ 的特征函数，$\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，则子样平均值 $\bar{\xi}$ 的特征函数为：

$$\varphi_1(t)=[\varphi_0(t)]^n=\left[\varphi\left(\frac{t}{n}\right)\right]^n, \qquad (6.2.9)$$

其中 $\varphi_0(t)=\varphi\left(\frac{t}{n}\right)$，它是$\frac{1}{n}\xi$的特征函数.

例 6.2.1 设总体 ξ 服从正态 $N(a, \sigma)$，求子样平均值 $\bar{\xi}$ 的分布.

解 因 ξ 的特征函数为

$$\varphi(t)=\exp\left[-\frac{\sigma^2}{2}t^2+jat\right],$$

则 $\bar{\xi}$ 的特征函数为

$$\varphi_1(t)=\left\{\exp\left[-\frac{\sigma^2}{2}\left(\frac{t}{n}\right)^2+ja\frac{t}{n}\right]\right\}^n$$
$$=\exp\left[-\frac{t^2}{2}\left(\frac{\sigma}{\sqrt{n}}\right)^2+jat\right],$$

即 $\bar{\xi}$ 服从正态 $N\left(a, \frac{\sigma}{\sqrt{n}}\right)$.

例 6.2.2 设总体 ξ 服从具有参数 λ 的泊松分布，求子样平均值 $\bar{\xi}$ 的分布.

解 因 ξ 的特征函数为

$$\varphi(t)=\exp[\lambda(e^{jt}-1)],$$

则 $\bar{\xi}$ 的特征函数为

$$\varphi_1(t)=\exp[n\lambda(e^{j\frac{t}{n}}-1)],$$

即 $\bar{\xi}$ 的概率分布为

$$P\left(\bar{\xi}=\frac{k}{n}\right)=e^{-n\lambda}\frac{(n\lambda)^k}{k!}, k=0,1,2,\cdots.$$

右端是参数为 $n\lambda$ 的泊松分布.

例 6.2.3 设总体 ξ 服从具有参数为 λ 的指数分布，求子样平均值 $\bar{\xi}$ 的分布.

$\bar{\xi}$ 的密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0, \lambda>0, \\ 0 \quad , & x\leqslant 0.\end{cases}$$

解 ξ的特征函数为

$$\varphi(t)=\frac{\lambda}{\lambda-jt}=\left(1-j\frac{t}{\lambda}\right)^{-1} \qquad \lambda>0,$$

则 $\bar{\xi}$ 的特征函数为

$$\varphi_{\bar{\xi}}(t)=\left(1-j\frac{t}{n\lambda}\right)^{-n}.$$

可见子样平均 $\bar{\xi}$ 服从具有参数 $\alpha=n-1, \beta=\frac{1}{n\lambda}$的 Γ- 分布，记作 $\Gamma\left(n-1, \frac{1}{n\lambda}\right)$.

应用第五章的定理 5.4.3，再举一例.

例 6.2.4 设总体 ξ 服从参数为 p 的二项分布，$0<p<1$，N 为正整数，概率函数为：

$$P\{\xi=k\}=C_N^k p^k(1-p)^{N-k} \qquad k=0,1,\cdots,N,$$

求子样平均值 $\bar{\xi}$ 的极限分布.

解 因 $E(\xi)=Np \qquad D(\xi)=Np(1-p)\leqslant\frac{N}{4}$

则统计量

$$\eta_n=\frac{\sum_{i=1}^{n}\xi_i-E\left(\sum_{i=1}^{n}\xi_i\right)}{\sqrt{D\left(\sum_{i=1}^{n}\xi_i\right)}}=\frac{n\bar{\xi}-nNp}{\sqrt{nNp(1-p)}}$$

$$=\frac{\bar{\xi}-Np}{\sqrt{Np(1-p)/n}}$$

有极限分布 $N(0,1)$，称 $\bar{\xi}$ 有渐近正态分布 $N\left(Np, \sqrt{\frac{Np(1-p)}{n}}\right)$

*2. 极值的分布

极值分布在水文、气象、地震等预报问题中，有其重要的作用．在大型工程的设计中，我们需要对一些带有严重破坏性的自然灾害进行必要的估计与预测．如在建造桥梁时，为了防止洪水冲塌桥梁这类事故发生，设计时就必须预先考虑到在使用期间该河流可能爆发的最高水位；在建水库、河堤时也有这样的问题．又如，在建造高大建筑物时，也要考虑到今后若干年内的最高水位、最大风压、地震的最大震级等．考察了解这些随机变数的概率分布，就是极值的分布问题．

设总体 ξ 的分布函数为 $F(x)$，$\xi_1,\cdots,\xi_n$ 为其子样，子样中的最大项 $\xi_n^*=\max\{\xi_1,\cdots,\xi_n\}$ 的分布函数记为 $F_n(u)$，则有：

$$\begin{aligned} F_n(u) &= P\{\xi_n^*<u\} \\ &= P\{\xi_1<u,\cdots,\xi_n<u\}=[F(u)]^n. \end{aligned} \qquad (6.2.10)$$

子样中的最小项 $\xi_1^*=\min\{\xi_1,\cdots,\xi_n\}$ 的分布函数记为 $F_1(v)$，由于

$$\begin{aligned} P\{\xi_1^*\geqslant v\} &= P\{\xi_1\geqslant v,\cdots,\xi_n\geqslant v\} \\ &= [P\{\xi\geqslant v\}]^n=[1-F(v)]^n, \end{aligned}$$

所以

$$F_1(v)=1-[1-F(v)]^n. \qquad (6.2.11)$$

如果总体 ξ 有分布密度函数 $f(x)$，则 ξ_n^* 及 ξ_1^* 的分布密度函数分别为

$$f_n(u)=nf(u)[F(u)]^{n-1}, \qquad (6.2.12)$$

$$f_1(v)=nf(v)[1-F(v)]^{n-1}. \qquad (6.2.13)$$

一般地，若 $\xi_1^*<\xi_2^*<\cdots<\xi_n^*$ 为子样 $\xi_1,\xi_2,\cdots,\xi_n$ 的顺序统计量，由小到大的第 k 位统计量为 ξ_k^*，其分布函数记为 $F_k(y)$，$k=1,2,\cdots,n$．我们用 $P\{y\leqslant\xi_k^*<y+\Delta y\}$ 表示事件"在子样 $\xi_1,\xi_2,\cdots,\xi_n$ 中有一个落在 $[y,y+\Delta y)$，$k-1$ 个落在 $(-\infty,y)$，其余 $n-k$ 个落在 $[y+\Delta y,\infty)$"的概率，于是

$$\Delta F_k(y)=P\{y\leqslant\xi_k^*<y+\Delta y\}$$

$$=C_n^1\Delta F(y)\cdot C_{n-1}^{k-1}[F(y)]^{k-1}\cdot[1-F(y+\Delta y)]^{n-k},$$

因此

$$F_k(y)=nC_{n-1}^{k-1}\int_{-\infty}^{y}[F(x)]^{k-1}\cdot[1-F(x)]^{n-k}dF(x). \quad (6.2.14)$$

特别地有

$$F_1(y)=1-[1-F(y)]^n,$$

$$F_n(y)=[F(y)]^n.$$

*3. 极差的分布

设总体 ξ 的分布函数为 $F(x)$，$\xi_1,\cdots,\xi_n$ 为其子样，子样极差 $D_n^*=\xi_n^*-\xi_1^*$ 的分布函数记为 $F_{D_n^*}(y)$，其中 $y\geqslant0$. 欲求 $F_{D_n^*}(y)$，一般地，可先导出顺序统计量 $(\xi_1^*,\cdots,\xi_n^*)$ 的联合分布，或者 (ξ_1^*,ξ_n^*) 的联合分布，从而导出子样极差 D_n^* 的分布函数. 现在，我们来直接导出 D_n^* 的分布函数 $F_{D_n^*}(x)$.

现对连续型随机变数 ξ 作讨论. 设总体 ξ 有密度函数 $f(x)$，$\xi_1,\cdots,\xi_n$ 为其子样，用 A 表示"子样 $\xi_1,\cdots,\xi_n$ 中有一个落在区间 $(v,v+dv)$ 而另 $n-1$ 个同时落在区间 $(v,v+y)$"这个事件，那么事件 A 发生的概率为

$$C_n^1f(v)dv\cdot[F(v+y)-F(v)]^{n-1},$$

其中 $(v,v+y)$ 为子样观察值的范围，y 为极差 D_n^* 的观察值的上界. 于是

$$F_{D_n^*}(y)=P\{D_n^*<y\}$$

$$=\int_{-\infty}^{\infty}n[F(v+y)-F(v)]^{n-1}f(v)dv. \quad (6.2.15)$$

其密度函数为

$$f_{D_n^*}(y)=\frac{dF_{D_n^*}(y)}{dy}$$

$$=n(n-1)\int_{-\infty}^{\infty}\left[\int_{v}^{v+y}f(x)dx\right]^{n-2}\cdot f(v+y)f(v)dv. \quad (6.2.16)$$

子样极差 D_n^* 的数学期望及方差分别为

$$c_n = E(D_n^*) = \int_0^\infty y f_{D_n^*}(y) dy, \tag{6.2.17}$$

$$v_n^2 = D(D_n^*) = \int_0^\infty (y - c_n)^2 f_{D_n^*}(y) dy. \tag{6.2.18}$$

例 6.2.5 设总体 ξ 服从参数为 λ 的指数分布，其密度函数及分布函数分别为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x < \infty, \lambda > 0, \\ 0, & \text{其它} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & 0 < x < \infty, \lambda > 0, \\ 0, & \text{其它} \end{cases}$$

求子样 $\xi_1, \cdots, \xi_n$ 的极值与极差的分布.

解 由(6.2.12)、(6.2.13)及(6.2.16)分别求出

$$f_n(u) = \begin{cases} n\lambda e^{-\lambda u}[1 - e^{-\lambda u}]^{n-1}, & 0 < u < \infty, \\ 0, & \text{其它} \end{cases} \tag{6.2.19}$$

$$f_1(v) = \begin{cases} n\lambda e^{-n\lambda v}, & 0 < v < \infty, \\ 0, & \text{其它} \end{cases} \tag{6.2.20}$$

$$f_{D_n^*}(y) = \begin{cases} (n-1)[1 - e^{-\lambda y}]^{n-2} \lambda e^{-\lambda y}, & 0 < y < \infty. \\ 0, & \text{其它} \end{cases} \tag{6.2.21}$$

可见，子样的最小值 ξ_1^* 服从参数为 $n\lambda$ 的指数分布.

在一般情况下，子样的极值与极差这些统计量的分布函数并不那么容易求出来. 例如当总体 ξ 服从正态 $N(a, \sigma)$ 或总体 ξ 服从分布 $\Gamma(\alpha, \beta)$，它们的子样极值的分布函数表达式都相当复杂. 格涅坚科在他的 1943 年的文章中具体地讨论了子样最大值 ξ_n^* 的极限分布的三种类型，并找到了向这些极限分布收敛的充要条件(参看[2]中的第十三章极值分布).

设总体 ξ 服从正态 $N(0,1)$，即有密度函数

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

对于子样容量的不同 n 值，利用(6.2.17)及(6.2.18)两式，可算出 c_n 及 v_n^2 的数值，如表 6.2.1 所示。

表 6.2.1

n	c_n	$1/c_n$	v_n
2	1.12838	0.8862	0.853
3	1.69257	0.5908	0.888
4	2.05875	0.4857	0.880
5	2.32593	0.4299	0.864
6	2.53441	0.3946	0.848
7	2.70436	0.3698	0.833
8	2.84720	0.3512	0.820
9	2.97003	0.3367	0.808
10	2.07751	0.3249	0.797

若总体 ξ 服从正态 $N(a,\sigma)$，$\xi_1,\cdots,\xi_n$ 为其子样，极差为

$$D_n^*=\max_{1\leq i\leq n}\xi_i-\min_{1\leq i\leq n}\xi_i. \tag{6.2.22}$$

作标准化变换

$$\xi'=\frac{\xi-a}{\sigma}$$

$$\xi_i'=\frac{\xi_i-a}{\sigma}\qquad i=1,\cdots,n,$$

则 ξ' 服从正态 $N(0,1)$，$\xi_1',\cdots,\xi_n'$ 为其子样，并有

$$\begin{aligned}\frac{D_n^*}{\sigma}&=\max_{1\leq i\leq n}\left(\frac{\xi_i-a}{\sigma}\right)-\min_{1\leq i\leq n}\left(\frac{\xi_i-a}{\sigma}\right)\\&=\max_{1\leq i\leq n}\xi_i'-\min_{1\leq i\leq n}\xi_i'=D_n^{*\prime},\end{aligned}$$

$$E\left(\frac{D_n^*}{\sigma}\right)=E(D_n^{*\prime})=c_n.$$

其中 c_n 是在总体 ξ' 服从正态 $N(0,1)$的假定下，由(6.2.17)式计算得到的(它在表(6.2.1)中可查得到)．于是

$$E\left(\frac{D_n^*}{c_n}\right)=\sigma, D(D_n^*)=v_n^2\sigma^2. \qquad (6.2.23)$$

在第七章的§7.2中将看到，统计量 D_n^*/c_n 是 σ 的无偏估计量.

§6.3 抽样分布定理

现在，我们来讨论几个重要的有关抽样分布的定理，这些定理在估计理论、假设检验及方差分析等数理统计学的基本内容中都有重要的作用．这些定理，都是在总体 ξ 服从正态 $N(a,\sigma)$的基本假定下而得到的．

定理 6.3.1 设总体 ξ 服从正态 $N(a,\sigma)$，ξ_1，…，ξ_n 为其子样，子样的平均值与方差分别记为 $\bar{\xi}$ 与 S^2，则 $\bar{\xi}$ 服从正态 $N\left(a,\frac{\sigma}{\sqrt{n}}\right)$，$\frac{nS^2}{\sigma^2}$服从自由度为 $n-1$ 的 χ^2-分布，简记作$\frac{nS^2}{\sigma^2}\sim\chi^2_{(n-1)}$，而且 $\bar{\xi}$ 与 S^2 相互独立

证 先就 $a=0$ 的情形证明所需要的结论．今对子样作正交线性变换如下：

$$\begin{cases}\eta_1=\dfrac{1}{\sqrt{1\cdot 2}}[\xi_1-\xi_2],\\ \eta_2=\dfrac{1}{\sqrt{2\cdot 3}}[\xi_1+\xi_2-2\xi_3],\\ \vdots \\ \eta_{n-1}=\dfrac{1}{\sqrt{(n-1)\cdot n}}[\xi_1+\xi_2+\cdots+\xi_{n-1}-(n-1)\xi_n],\\ \eta_n=\dfrac{1}{\sqrt{n}}[\xi_1+\xi_2+\cdots+\xi_n].\end{cases} \qquad (6.3.1)$$

我们来证 $\eta_1, \cdots, \eta_n$ 相互独立且有相同正态 $N(0,\sigma)$.

(1) 由于正交变换保持长度不变，所以

$$\sum_{i=1}^{n} \eta_i^2 = \sum_{i=1}^{n} \xi_i^2. \tag{6.3.2}$$

(2) 注意到$(\xi_1, \cdots, \xi_n)$的联合分布是 n 维正态的，其联合密度函数为

$$f(x_1, \cdots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{x_i^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2}. \tag{6.3.3}$$

由于(6.3.1)变换的雅可比行列式的绝对值为 1，所以$(\eta_1, \cdots, \eta_n)$的联合密度函数为

$$g(y_1, \cdots, y_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\frac{2}{2\sigma^2}\sum_{i=1}^{n} y_i^2}. \tag{6.3.4}$$

于是$(\eta_1, \cdots, \eta_n)$的联合分布是 n 维正态分布. 我们知道，联合正态分布的边沿分布也是正态的，且在正态情形，不相关性与独立性是等价的，因而 $\eta_1, \cdots, \eta_n$ 是相互独立且有相同正态 $N(0,\sigma)$的随机变数.

(3) 在(6.3.1)式中， $\eta_n = \frac{n}{\sqrt{n}}\bar{\xi} = \sqrt{n}\,\bar{\xi}$,

即

$$\eta_n^2 = n\bar{\xi}^2$$

则

$$nS^2 = \sum_{i=1}^{n} (\xi_i - \bar{\xi})^2 = \sum_{i=1}^{n} \xi_i^2 - n\bar{\xi}^2$$

$$= \sum_{i=1}^{n} \eta_i^2 - \eta_n^2 = \sum_{i=1}^{n-1} \eta_i^2,$$

于是

$$\frac{nS^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n-1}\eta_i^2\sim\chi^2_{(n-1)}. \tag{6.3.5}$$

而且 nS^2 与 η_n 相互独立，也就是 nS^2 与 $\bar{\xi}$ 相互独立，这就证明了我们所需的结论。

除去 $a=0$ 这个假定，定理的结论仍然成立。事实上，我们可作变换：

$$\zeta=\xi-a,\quad \zeta_i=\xi_i-a,\quad i=1,\cdots,n. \tag{6.3.6}$$

由 ξ 服从正态 $N(a,\sigma)$ 知，ζ 服从正态 $N(0,\sigma)$。另外

$$\bar{\zeta}=\bar{\xi}-a,$$

$$\frac{1}{n}\sum_{i=1}^{n}(\zeta_i-\bar{\zeta})^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=S^2,$$

对 ζ 应用上述证明的结果可知，$\sum_{i=1}^{n}(\zeta_i-\bar{\zeta})^2$ 与 $\bar{\zeta}$ 相互独立，从而 nS^2 与 $\bar{\xi}$ 相互独立；而且 $\frac{nS^2}{\sigma^2}$ 服从自由度为 $n-1$ 的 χ^2-分布，即 $\frac{nS^2}{\sigma^2}\sim\chi^2_{(n-1)}$.■

推论 设总体 ξ 服从正态 $N(a,\sigma)$，$\xi_1,\cdots,\xi_n$ 为子样，则统计量

$$T=\sqrt{n-1}\,\frac{\bar{\xi}-a}{S} \tag{6.3.7}$$

服从自由度为 $n-1$ 的 t-分布，记作 $T\sim t_{(n-1)}$

证 由定理 6.3.1 知，下述两个随机变数

$$\zeta=\frac{\bar{\xi}-a}{\sigma/\sqrt{n}},$$

$$\eta=\frac{nS^2}{\sigma^2} \tag{6.3.8}$$

相互独立，而且 ξ 服从正态 $N(0,1)$，η 服从自由度为 $n-1$ 的 χ^2-分布，由第二章的§2.4知，统计量

$$\frac{\xi}{\sqrt{\eta/(n-1)}}=\sqrt{n-1}\frac{\bar{\xi}-a}{S}$$

服从自由度为 $n-1$ 的 t-分布，记作 $\sqrt{n-1}\frac{\bar{\xi}-a}{S}\sim t_{(n-1)}$. ∎

定理 6.3.2 设总体 ξ 服从正态 $N(a_1,\sigma)$，$\xi_1,\cdots,\xi_{n_1}$ 为其子样，总体 η 服从正态 $N(a_2,\sigma)$，$\eta_1,\cdots,\eta_{n_2}$ 为其子样，而且这两个子样是相互独立的①. 记

$$\bar{\xi}=\frac{1}{n_1}\sum_{i=1}^{n_1}\xi_i,\qquad S_1^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(\xi_i-\bar{\xi})^2,$$

$$\bar{\eta}=\frac{1}{n_2}\sum_{j=1}^{n_2}\eta_j,\qquad S_2^2=\frac{1}{n_2}\sum_{j=1}^{n_2}(\eta_j-\bar{\eta})^2,$$

则有：

(1) $$\frac{(n_2-1)n_1S_1^2}{(n_1-1)n_2S_2^2}\sim F_{(n_1-1,n_2-1)},\tag{6.3.9}$$

即左端的统计量服从第一自由度为 n_1-1、第二自由度为 n_2-1 的 F-分布.

(2) $$\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sqrt{n_1S_1^2+n_2S_2^2}}\sim t_{(n_1+n_2-2)}.\tag{6.3.10}$$

即左端的统计量服从自由度为 n_1+n_2-2 的 t-分布.

证 先证(6.3.9)式是正确的. 由定理 6.3.1 知

$$\frac{n_1S_1^2}{\sigma^2}\sim\chi^2_{(n_1-1)},$$

$$\frac{n_2S_2^2}{\sigma^2}\sim\chi^2_{(n_2-1)}.$$

---

① 随机向量 $(\xi_1,\cdots,\xi_{n_1})$ 与随机向量 $(\eta_1,\cdots,\eta_{n_2})$ 之间是相互独立的.

由于两个子样是相互独立的，因而统计量$\frac{n_1S_1^2}{\sigma^2}$与统计量$\frac{n_2S_2^2}{\sigma^2}$也相互独立. 由第二章§2.4知

$$\frac{\left(\frac{n_1S_1^2}{\sigma^2}\right)\Big/(n_1-1)}{\left(\frac{n_2S_2^2}{\sigma^2}\right)\Big/(n_2-1)}=\frac{(n_2-1)n_1S_1^2}{(n_1-1)n_2S_2^2}\sim F_{(n_1-1,n_2-1)},$$

即左端的统计量服从第一自由度为n_1-1，第二自由度为n_2-1的F分布.

现证(6.3.10)式是正确的. 由于两个子样是相互独立的，因而统计量$\bar{\xi}$与$\bar{\eta}$相互独立，而且

$$E(\bar{\xi}-\bar{\eta})=a_1-a_2,$$

$$D(\bar{\xi}-\bar{\eta})=\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}.$$

容易证明统计量

$$(\bar{\xi}-a_1)-(\bar{\eta}-a_2)=(\bar{\xi}-\bar{\eta})-(a_1-a_2)$$

服从正态$N\left(0,\ \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$. 由于两个子样是相互独立的，且

$\frac{n_1S_1^2}{\sigma^2}\sim\chi^2_{(n_1-1)}$，$\frac{n_2S_2^2}{\sigma^2}\sim\chi^2_{(n_2-1)}$. 由第二章§2.4知，

$$\frac{n_1S_1^2}{\sigma^2}+\frac{n_2S_2^2}{\sigma^2}=\frac{n_1S_1^2+n_2S_2^2}{\sigma^2}\sim\chi^2_{(n_1+n_2-2)}.$$

如果能证明统计量$(\bar{\xi}-\bar{\eta})-(a_1-a_2)$与$(n_1S_1^2+n_2S_2^2)/\sigma^2$是相互独立的，则由第二章§2.4知，统计量

$$\frac{[(\bar{\xi}-\bar{\eta})-(a_1-a_2)]\Big/\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\sqrt{\frac{n_1S_1^2+n_2S_2^2}{\sigma^2}\Big/(n_1+n_2-2)}}$$

$$=\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sqrt{n_1S_1^2+n_2S_2^2}}\sim t_{(n_1+n_2-2)}.$$

下面证明$(\xi-a_1)-(\eta-a_2)$与$(n_1S_1^2+n_2S_2^2)/\sigma^2$是相互独立的.由定理 6.3.1 知道,$(\xi-a_1)$与$n_1S_1^2/\sigma^2$相互独立,$(\eta-a_2)$与$n_2S_2^2/\sigma^2$相互独立.由于两个子样是相互独立的,因而$(\xi-a_1)$与$n_2S_2^2/\sigma^2$相互独立,$(\eta-a_2)$与$n_1S_1^2/\sigma^2$相互独立,于是$(\xi-a_1)-(\eta-a_2)$与$n_1S_1^2/\sigma^2$相互独立,并与$n_2S_2^2/\sigma^2$相互独立,因而与$(n_1S_1^2+n_2S_2^2)/\sigma^2$相互独立 ∎

利用定理 6.3.3,对这个结论可作严格的证明(留作练习).

定理 6.3.3(柯赫伦分解定理) 设$\xi_1,\cdots,\xi_n$为n个相互独立的随机变数,每一个$\xi_i(i=1,\cdots,n)$都服从正态$N(0,1)$.如果

$$Q=\sum_{i=1}^{n}\xi_i^2=\sum_{l=1}^{k}Q_l,$$

其中Q_l是秩为n_l的$\xi_1,\cdots,\xi_n$的二次型,则下述结论

(1) $Q_1,\cdots,Q_k$相互独立,

(2) $Q_l\sim\chi^2_{n_l}$ $\quad l=1,\cdots,k$

成立的充要条件为$\sum\limits_{l=1}^{k}n_l=n$.

证 必要性 设$Q_1,\cdots,Q_k$相互独立且分别服从自由度为$n_1,\cdots,n_k$的χ^2-分布,即

$$Q_l\sim\chi^2_{n_l}\qquad(l=1,\cdots,k),$$

则由定理 2.4.2 知

$$\sum_{l=1}^{k}Q_l\sim\chi^2_{\sum\limits_{l=1}^{k}n_l}.$$

由定理的假设知

$$Q=\sum_{i=1}^{n}\xi_i^2\sim\chi^2_n,$$

且

$$Q=\sum_{i=1}^{n}\xi_i^2=\sum_{l=1}^{k}Q_l,\tag{6.3.11}$$

所以

$$n=\sum_{l=1}^{k} n_l.$$

充分性 若 $n=\sum_{l=1}^{k} n_l$，我们来证(1)及(2)成立．由于 Q_l 是秩为 n_l 的 $\xi_1,\cdots,\xi_n$ 的二次型，可以证明(留作练习)Q_l 能够表成下式：

$$Q_l=\sum_{j=1}^{n_l} b_{lj}\eta_{lj}^2, \qquad b_{lj}=1 \text{ 或 } -1, \tag{6.3.12}$$

其中 η_{lj} 是 $\xi_1,\cdots,\xi_n$ 的线性函数($j=1,\cdots,n_l$)．于是

$$Q=\sum_{i=1}^{n}\xi_i^2=\sum_{l=1}^{k}\sum_{j=1}^{n_l} b_{lj}\eta_{lj}^2. \tag{6.3.13}$$

记

$$\mathbf{x}'=(\xi_1,\cdots,\xi_n),$$

$$\mathbf{y}'=(\eta_{11},\cdots,\eta_{1n_1},\cdots,\eta_{k1},\cdots,\eta_{kn_k}).$$

则(6.3.13)式可以写为矩阵形式：

$$\mathbf{x}'\mathbf{x}=\mathbf{y}'\mathbf{B}\mathbf{y}.$$

其中$\mathbf{B}$ 是对角线元素为 ± 1 的 $n\times n$ 的对角矩阵，$n=\sum_{l=1}^{k} n_l$．作变换

$$\mathbf{y}=\mathbf{A}\mathbf{x},$$

它是由正定二次型$\mathbf{x}'\mathbf{x}$ 变到 $\mathbf{y}'\mathbf{B}\mathbf{y}$ 的变换式．由

$$\mathbf{x}'\mathbf{x}=\mathbf{y}'\mathbf{B}\mathbf{y}=\mathbf{x}'\mathbf{A}'\mathbf{B}\mathbf{A}\mathbf{x}$$

得到 $\mathbf{A}'\mathbf{B}\mathbf{A}=\mathbf{I}$，由此推知 $\mathbf{A}$ 是非奇异的，于是$\mathbf{A}\mathbf{A}'$ 是正定的．因此，$\mathbf{B}=\mathbf{I}$ 及 $\mathbf{A}'\mathbf{A}=\mathbf{I}$．从而 $\mathbf{A}$ 为正交矩阵，由定理6.3.1知，η_{lj} 之间相互独立而且都服从正态 $N(0,1)$，($l=1,\cdots,k,j=1,\cdots,n_l$)．所以$Q_l\sim\chi^2_{n_l}$，而且 $Q_1,\cdots,Q_k$ 是相互独立的．▌

推论 设$\xi_1, \cdots, \xi_k$为k个相互独立的且每个都服从正态$N(0,1)$的随机变数，今从这k个总体中分别抽取容量为n_i的子样$\xi_{i1}, \cdots, \xi_{in_i}$ $(i=1, \cdots, k)$，记

$$\bar{\xi}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\xi_{ij},$$

$$S_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)^2,$$

$$n = \sum_{i=1}^{k} n_i,$$

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\xi_{ij} = \frac{1}{n}\sum_{i=1}^{k} n_i\bar{\xi}_i.$$

则

$$(1)\quad S_总^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi})^2 = \sum_{i=1}^{k} n_i S_i^2 + \sum_{i=1}^{k} n_i(\bar{\xi}_i-\bar{\xi})^2; \tag{6.3.14}$$

(2) 记 $Q_e = \sum_{i=1}^{k} n_i S_i^2,$

$$U = \sum_{i=1}^{k} n_i(\bar{\xi}_i-\bar{\xi})^2,$$

那么

$$\frac{U/(k-1)}{Q_e/(n-k)} \sim F_{(k-1, n-k)}. \tag{6.3.15}$$

证 先验证恒等式(6.3.14)，称(6.3.14)式为离差平方和的分解式.

$$S_总^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi})^2$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i}[\xi_{ij}-\bar{\xi}_i+\bar{\xi}_i-\bar{\xi}]^2$$

$$=\sum_{i=1}^{k}\sum_{j=1}^{n_i}[(\bar{\xi}_i-\bar{\xi})^2+2(\bar{\xi}_i-\bar{\xi})(\xi_{ij}-\bar{\xi}_i)+(\xi_{ij}-\bar{\xi}_i)^2]$$

$$=\sum_{i=1}^{k}n_i(\bar{\xi}_i-\bar{\xi})^2+\sum_{i=1}^{k}n_iS_i^2$$

$$=U+Q_e,$$

其中

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{\xi}_i-\bar{\xi})(\xi_{ij}-\bar{\xi}_i)$$

$$=\sum_{i=1}^{k}(\bar{\xi}_i-\bar{\xi})\left[\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)\right]=0.$$

现证(2). 由定理 6.3.1 可知, $S_k^2\sim\chi^2_{(n-1)}$. 我们知道, 在 Q_e 中有 k 个独立的线性关系式, 即

$$\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)=0\quad(i=1,\cdots,k),$$

Q_e 的自由度为 $\sum_{i=1}^{k}n_i-k=n-k$. 在 U 中有一个线性关系式, 即

$$\sum_{i=1}^{k}n_i(\bar{\xi}_i-\bar{\xi})=0.$$

因而 U 的自由度为 $k-1$. 于是由定理 6.3.3 知, Q_e 与 U 相互独立且有 $Q_e\sim\chi^2_{(n-k)}$ 及 $U\sim\chi^2_{(k-1)}$, 由定理 2.4.5 知

$$\frac{U/(k-1)}{Q_e/(n-k)}\sim F_{(k-1,n-k)}.\ \blacksquare$$

习　题

1. 设总体 ξ 服从正态 $N(12,2)$，今抽取容量为 5 的子样 $\xi_1,\cdots,\xi_5$，试问：

(1) 子样的平均值 $\bar{\xi}$ 大于 13 的概率是多少？

(2) 子样的极小值小于 10 的概率是多少？

(3) 子样的极大值大于 15 的概率是多少？

2. 设总体 ξ 服从几何分布，即 $P\{\xi=l\}=q^l p$, $l=0,1,2,\cdots$. 其中 $0<p<1$, $q=1-p$. 求 $M=\max\{\xi_1,\cdots,\xi_n\}$, $K=\min\{\xi_1,\cdots,\xi_n\}$ 的概率分布，$\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样.

3. 设电子元件的寿命(时数) ξ 服从以 $\lambda=0.0015$ 为参数的指数分布，即有密度函数 $f(x)=0.0015e^{-0.0015x}$, $x>0$. 今测试 6 个元件，并记录下它们各自失效的时间(单位：小时). 试问：

(1) 至 800 小时时没有一个元件失效的概率是多少？

(2) 至 3000 小时时所有元件都失效的概率是多少？

4. 设总体 ξ 服从泊松分布 $P(x,\lambda)$ 时，问子样的平均值 $\bar{\xi}$ 的渐近分布是什么？

5. 设 $\xi_1,\cdots,\xi_n$ 为总体 $\xi\sim N(a,\sigma)$ 的子样，$E(\xi)=a$, $D(\xi)=\sigma^2$，定义

$$d=\frac{1}{n}\sum_{i=1}^{n}|\xi_i-a|,$$

试证 $E(d)=\sqrt{\frac{2}{\pi}}\sigma$, $D(d)=\left(1-\frac{2}{\pi}\right)\frac{\sigma^2}{n}$.

6. 设总体 ξ 服从正态 $N(20,\sqrt{3})$，今从中抽取容量为 10 及 15 的两个独立子样，试问这两个子样的平均值之差的绝对值大于 0.3 的概率是多少？

7. 设总体 ξ 服从正态 $N(a,\sigma)$, ξ_1,ξ_2 为其子样，试求子样极差的分布，极大值与极小值的分布.

8. 设总体 ξ 服从参数为 λ 的指数分布，ξ_1,ξ_2 为其子样，试求子样的极大值、极小值与极差的分布.

9. 设总体 ξ 服从正态 $N(a_1,\sigma_1)$，总体 η 服从正态 $N(a_2,\sigma_2)$，$\bar{\xi}$ 及 $\bar{\eta}$，S_1^2 及 S_2^2 分别为其子样的平均值与方差，这两个子样的相关系数为：

$$R=\frac{\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\eta_i-\bar{\eta})}{S_1\cdot S_2}=\frac{S_{12}}{S_1\cdot S_2},$$

试证当(ξ_1,η_1)，…，(ξ_n,η_n)为正态总体(ξ,η)的子样时，则有

$$\sqrt{n-1}\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sqrt{S_1^2+S_2^2-2RS_1S_2}}\sim t_{(n-1)}.$$

10. 设总体ξ的$E(\xi)=a_1$，$D(\xi)=\sigma^2$，总体η的$E(\eta)=a_2$，$D(\eta)=\sigma^2$。n_1及n_2、$\bar{\xi}$及$\bar{\eta}$、S_1^2及S_2^2分别为其子样的容量大小、平均值与方差. 试证：当ξ及η服从正态分布且两个子样相互独立，则有

$$\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sqrt{S_1^2/n_2+S_2^2/n_1}}\underset{\substack{n_1\to\infty\\ n_2\to\infty}}{\sim}N(0,1).$$

11. 试证(6.3.12)式成立.

12. 设$\xi_1,\cdots,\xi_n$是n个相互独立的且都是服从正态$N(0,1)$的随机变数，$\xi_1,\cdots,\xi_n$到$\eta_1,\cdots,\eta_n$的变换为正交变换. 试证$\eta_1,\cdots,\eta_n$是n个相互独立的且都是服从正态$N(0,1)$的随机变数.

13. 设总体ξ服从正态$N(a,\sigma)$，$\xi_1,\cdots,\xi_n$为其子样，$\bar{\xi}$及S_n^2分别为子样的平均值及方差. 又设ξ_{n+1}服从正态$N(a,\sigma)$，且与$\xi_1,\cdots,\xi_n$相互独立. 试求统计量

$$\eta=\frac{\xi_{n+1}-\bar{\xi}}{S_n}\sqrt{\frac{n-1}{n+1}}$$

的抽样分布。

14. 设$\xi_1,\cdots,\xi_n$相互独立且分别服从正态$N(a_i,\sigma_i)$，试证$\eta=\sum_{i=1}^{n}c_i\xi_i$服从正态$N\left(\sum_{i=1}^{n}c_ia_i,\sqrt{\sum_{i=1}^{n}c_i^2\sigma_i^2}\right)$.

15. 设$\xi_1,\cdots,\xi_n$相互独立且分别服从正态$N(a,\sigma_i)$，$i=1,\cdots,n$，即数学期望相同而方差不同. 试证$\eta=\sum_{i=1}^{n}\frac{\xi_i}{\sigma_i}\Big/\sum_{i=1}^{n}\frac{1}{\sigma_i}$与$\zeta=\sum_{i=1}^{n}\left[\frac{\xi_i-a}{\sigma_i}-\frac{\eta-a}{n}\sum_{i=1}^{n}\left(\frac{1}{\sigma_i}\right)\right]^2$相互独立，而且$\eta$服从正态分布，$\zeta\sim\chi^2_{(n-1)}$.

16. 设总体ξ在$\left(\theta-\frac{1}{2},\theta+\frac{1}{2}\right)$上服从均匀分布，$\xi_1,\cdots,\xi_n$为其子样，$\xi_1^*,\cdots,\xi_n^*$为顺序统计量. 试求$\xi_1^*$，$\xi_n^*$及$(\xi_1^*,\xi_n^*)$的分布.

17. 试用定理6.3.3的结论，证明(6.3.10)式成立.

第七章 估计理论

估计理论是数理统计学的重要内容之一．我们知道，服从泊松分布的随机变数，其概率分布 $P(x;\lambda)$ 由一个参数 $\lambda>0$ 确定；服从正态分布的随机变数，其概率分布 $N(a,\ \sigma)$ 由一对参数 a 及 $\sigma>0$ 确定．就是说，对所要研究的随机变数 ξ，当它的概率分布的类型为已知时，还需确定分布函数中的参数是什么值，这样随机变数 ξ 的分布函数才能完全确定．例如，纺织厂细纱机上的断头次数 ξ 服从泊松分布，如果要求知道每只纱绽在某一时间间隔内断头次数为 k 次的概率是多少，那么就要首先确定泊松分布中的参数 λ 取什么值，然后才能得出所要求的概率．这就提出了参数的估计问题．设总体 ξ 的分布函数为 $F(x;\theta)$，其中参数 θ 为未知，试问怎样由子样 $\xi_1,\cdots,\xi_n$ 提供的信息，建立子样的函数即统计量来对未知参数 θ 作出估计，并且估计量的"最佳"准则又怎样确定？这类统计问题，称为参数的估计问题．另外在有些实际问题中，事先并不知道随机变数 ξ 服从什么分布，而要对其数字特征，如数学期望 $E(\xi)$ 及方差 $D(\xi)$ 等作出估计．例如，某砖瓦厂生产了一批青砖，为了评定这批青砖的质量情况，需要了解这批青砖的平均抗压强度．一般地说抗压强度为一随机变数，记作 ξ，现在的问题是希望了解 $E(\xi)$ 是多大．我们知道，随机变数 ξ 的数字特征同它的概率分布中的参数有一定关系，因而对数字特征的估计问题，也称之为参数的估计问题．

参数估计有点估计和区间估计两方面的问题．参数点估计有许多方法，在这一章讨论矩法、极大似然法及贝叶斯估计，在第九

章中将讨论最小二乘法估计．估计理论的另一方面是非参数性的估计问题，如随机变数 ξ 与 η 之间有一定的相关关系，试问在什么准则下，由一个随机变数对另一随机变数所作估计与预测为"最佳"．

§7.1 矩法与极大似然法

设总体 ξ 的分布函数为 $F(x;\theta)$，其中 θ 为未知参数，θ 的取值范围记为 Ω，Ω 称为参数空间①．例如，若随机变数 ξ 服从二项分布 $b(x;p)$，p 为参数，则 $p\in\Omega=(0,1)$；若 ξ 服从泊松分布 $P(x;\lambda)$，λ 为参数，则 $\lambda\in\Omega=(0,\infty)$；若 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是参数，则 $\Omega=\{(a,\sigma);a\in R_1=(-\infty,\infty),\sigma\in(0,\infty),\}$

定义 7.1.1 （点估计）设总体 ξ 的分布函数 $F(x;\theta)$ 中参数 θ 未知，$\theta\in\Omega$，Ω 为参数空间，今由子样 $\xi_1,\cdots,\xi_n$ 建立统计量 $T(\xi_1,\cdots,\xi_n)$，对于子样观察值 $(x_1,\cdots,x_n)$，若将 $T(x_1,\cdots,x_n)=t$ 作为 θ 的估计值，则称 $T(\xi_1,\cdots,\xi_n)$ 为 θ 的估计量，通常记作 $\hat{\theta}=T(\xi_1,\cdots,\xi_n)$．建立一个这样的统计量 $T(\xi_1,\cdots,\xi_n)$ 作为 θ 的估计量，称之为参数 θ 的点估计．

如果随机变数 ξ 的分布函数 $F(x;\theta_1,\cdots,\theta_l)$ 中有 l 个不同的未知参数，则要由子样 $\xi_1,\cdots,\xi_n$ 建立 l 个统计量作为这 l 个未知参数的估计量，分别记之为 $\hat{\theta}_1,\cdots,\hat{\theta}_l$．

现在叙述求参数点估计的矩法与极大似然法．

一、矩法

皮尔逊所引入的矩法是较早提出的求参数点估计的方法．我们从欣钦大数定律知道，若总体 ξ 的数学期望 $E(\xi)$ 有限，则子样

① 在第一章中我们以 Ω 表示样本空间；在数理统计各章中，我们用 Ω 表示参数空间．

的平均值 $\bar{\xi}$ 依概率收敛于 $E(\xi)$. 这就启发我们想到，在利用子样所提供的信息来对总体 ξ 的分布函数中未知参数 θ 作估计时，可以用子样矩作为总体矩的估计.

定义 7.1.2(矩法估计量) 设总体 ξ 的分布函数 $F(x;\theta_1,\cdots,\theta_l)$ 中有 l 个未知参数 $\theta_1,\cdots,\theta_l$，假定总体 ξ 的 l 阶原点矩 $E(\xi^l)$ 存在，

并记

$$\nu_k(\theta_1,\cdots,\theta_l)=E_{\theta_1,\cdots,\theta_l}(\xi^k)$$

$(k=1,2,\cdots,l)$. 由下列方程组

$$\begin{cases}\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i=\nu_1(\theta_1,\cdots,\theta_l),\\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2=\nu_2(\theta_1,\cdots,\theta_l),\\ \qquad\vdots\\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^l=\nu_l(\theta_1,\cdots,\theta_l)\end{cases}\tag{7.1.1}$$

解得 $\hat{\theta}_k=\hat{\theta}_k(\xi_1,\cdots,\xi_n)$ $(k=1,2,\cdots,l)$，并以 $\hat{\theta}_k$ 作为参数 θ_k 的估计量，则称 $\hat{\theta}_k$ 为未知参数 θ_k 的<u>矩法估计量</u>.

若 $\hat{\theta}$ 为 θ 的矩法估计量，$g(\theta)$ 为 θ 的连续函数，则也称 $g(\hat{\theta})$ 为 $g(\theta)$ 的矩法估计量. 矩法估计更一般的提法是，利用子样的数字特征作为总体的数字特征的估计. 例如，不论总体 ξ 服从什么分布，若 $E(\xi)=a$ 及 $D(\xi)=\sigma^2$ 是有限的，但其值未知，则由

$$\begin{cases}\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i=\nu_1,\\ \dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2=\nu_2,\end{cases}$$

解得：

$$\hat{a}=\hat{\nu}_1=\bar{\xi},$$

$$\hat{\sigma}^2=\hat{\nu}_2-(\hat{\nu}_1)^2=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-(\bar{\xi})^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=S^2. \qquad (7.1.2)$$

就是说，子样的平均值 $\bar{\xi}$ 及方差 S^2 分别为总体的数学期望 a 及方差 σ^2 的矩法估计量.

例 7.1.1　求事件概率 p 的矩法估计量

记事件 A 的概率 $P(A)=p$，今要求 p 的估计量. 我们用 ξ 表示事件 A 在一次试验中是否出现这样的随机变数，即：

$$\xi=\begin{cases}1, & A\text{出现},\\ 0, & A\text{不出现},\end{cases}$$

则 $P\{\xi=1\}=P(A)=p$, $P\{\xi=0\}=1-p$. 因为 $E(\xi)=p$，所以 p 的矩法估计量为：

$$\hat{p}=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\frac{\mu}{n}, \qquad (7.1.3)$$

其中 μ 为事件 A 在 n 次独立试验中出现的次数. 也就是说，在 n 次独立试验中，用事件 A 出现的频率 $\frac{\mu}{n}$ 作为事件 A 出现的概率 p 的矩法估计量.

例 7.1.2　设总体 ξ 在 $[\theta_1,\theta_2]$ 上服从均匀分布，其密度函数为：

$$f(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1}, & x\in[\theta_1,\theta_2],\\ 0, & x\bar{\in}[\theta_1,\theta_2],\end{cases}$$

其中 $\theta_2>\theta_1$，试求 θ_1 及 θ_2 的矩法估计量.

解　我们知道，$E(\xi)=\frac{1}{2}(\theta_1+\theta_2)$, $D(\xi)=\frac{1}{12}(\theta_1-\theta_2)^2$. 由

方程组

$$\begin{cases}\bar{\xi}=\frac{1}{2}(\theta_1+\theta_2)\\ S^2=\frac{1}{12}(\theta_1-\theta_2)^2\end{cases}$$

解得:

$$\begin{cases}\hat{\theta}_1=\bar{\xi}-\sqrt{3}S,\\ \hat{\theta}_2=\bar{\xi}+\sqrt{3}S,\end{cases} \tag{7.1.4}$$

则 $\hat{\theta}_1$、$\hat{\theta}_2$ 分别为 θ_1 及 θ_2 的矩法估计量.

例 7.1.3 设总体 ξ 服从 Γ-分布 $\Gamma(\alpha,\beta)$，其密度函数为:

$$f(x)=\begin{cases}\frac{\beta^{-(\alpha+1)}}{\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}, & x>0,\\ 0, & x\leqslant 0,\end{cases} \tag{7.1.5}$$

试求 α 及 β 的矩法估计量.

解 我们知道

$$E(\xi)=\beta(\alpha+1),\ D(\xi)=\beta^2(\alpha+1),$$

由方程组

$$\begin{cases}\bar{\xi}=\beta(\alpha+1)\\ S^2=\beta^2(\alpha+1)\end{cases}$$

解得

$$\hat{\alpha}=\frac{(\bar{\xi})^2}{S^2}-1,\ \hat{\beta}=\frac{S^2}{\bar{\xi}}. \tag{7.1.6}$$

例 7.1.4 求相关系数 ρ 的矩法估计量

解 记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\ S_1^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2,$$

$$\bar{\eta}=\frac{1}{n}\sum_{i=1}^{n}\eta_i,\ S_2^2=\frac{1}{n}\sum_{i=1}^{n}(\eta_i-\bar{\eta})^2.$$

用(6.2.13)所定义的子样相关系数R作为ρ的矩法估计量，即

$$\hat{\rho}=R=\frac{S_{12}}{S_1\cdot S_2}.$$

这可理解为：用$\bar{\xi}$及$\bar{\eta}$分别作为$E(\xi)$及$E(\eta)$的矩法估计量，用S_1^2及S_2^2分别作为$D(\xi)$及$D(\eta)$的矩法估计量，用子样协方差S_{12}作为总体的协方差$\mathrm{cov}(\xi,\eta)=E[\xi-E(\xi)][\eta-E(\eta)]$的矩法估计量，因而可用$R$作为$\rho$的矩法估计量，其中$\rho$为总体的相关系数

$$\rho=\frac{\mathrm{cov}(\xi,\eta)}{\sqrt{D(\xi)}\sqrt{D(\eta)}}.$$

矩法估计是最古老的点估计方法，它直观而又简便，特别是在对总体ξ的数学期望及方差数字特征作估计时，并不一定要知道ξ的分布函数$F(x;\theta)$．但是，矩法估计要求随机变数ξ的原点矩存在，如哥西分布的原点矩不存在，那就不能用矩法了．再者，子样矩的表达式同总体ξ的分布函数$F(x;\theta)$的表达式无关，因而矩法还没有充分利用$F(x;\theta)$对参数θ所提供的信息．

二、极大似然法

费歇引进的极大似然法，从理论观点来看，至今仍然是参数点估计中最重要的方法．从后面将会知道，这种估计方法，是利用总体ξ的分布函数$F(x;\theta)$的表达式及子样所提供的信息，建立未知参数θ的估计量$\hat{\theta}(\xi_1,\cdots,\xi_n)$．

极大似然法的想法同矩法一样也是直观的．今举一个通俗的例子，有两位同学一起进行实弹射击，两人共同射击一个目标，事先并不知道谁的技术较好，让每人各打一发，有一人击中目标，那么认为击中的同学的技术比击不中的技术较好，显然是合理的．又举一例：有一事件，我们知道它发生的概率p只可能是0.01或0.09，在一次观察中这事件发生了，试问这事件发生的概率是什么？当

缘人们会认为它发生的概率是 0.09 而不是 0.01. 我们再举一个实例，由此例便于引进极大似然法.

例 7.1.5 设在一个布袋中盛放着许多个白球和黑球，但不知道是黑球多还是白球多，只知道两种球的数目之比为 1:3，就是说抽取到黑球的概率是$\frac{1}{4}$或$\frac{3}{4}$，我们希望通过实验来判断黑球占的比例是$\frac{1}{4}$或$\frac{3}{4}$.

解 今用有放回抽取的方法从布袋中抽取 n 个球，其中黑球的个数记为 ξ，则 ξ 服从二项分布

$$P\{\xi=x\}=C_n^x p^x(1-p)^{n-x}. \tag{7.1.7}$$

今就 $n=3$ 的情形讨论如下. 怎样通过子样的观察值也即 x 的取值来估计参数 p 呢？换句话说，在什么情况下取 $p=\frac{1}{4}$而在另外的情况下取 $p=\frac{3}{4}$更为合理呢？为此，我们就 $p=\frac{1}{4}$或$\frac{3}{4}$为参数值的二项分布计算得到的概率列如下表 7.1.1.

表 7.1.1

x	0	1	2	3
$P\left(x;\frac{3}{4}\right)$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$P\left(x;\frac{1}{4}\right)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

我们从对子样所下的定义中知道，子样来自总体并很好地反映了总体的概率分布特征. 因此，我们在对总体的分布函数的参数作估计时，应该从子样的观察值出发来考虑. 在这个例子中，如果我们观察到的黑球个数 $x=0$，由 $P\left(0,\ \frac{1}{4}\right)=\frac{27}{64}$及 $P\left(0,\ \frac{3}{4}\right)=\frac{1}{64}$

知，显然$P\left(0,\frac{1}{4}\right)>P\left(0,\frac{3}{4}\right)$，这表明使$x=0$的子样从$p=\frac{1}{4}$为参数的总体中抽取比从参数 $p=\frac{3}{4}$的总体中抽取更有可能发生，因而取$\frac{1}{4}$作为 p 的估计比取$\frac{3}{4}$作为 p 的估计更为合理. 类似地，当$x=1$时，取$\frac{1}{4}$作为 p 的估计比取$\frac{3}{4}$更为合理. 而当 $x=2$ 或 3 时，取$\frac{3}{4}$作为 p 的估计比取$\frac{1}{4}$更合理. 综上所述，确定参数 p 的估计量为:

$$\hat{p}(x)=\begin{cases}\frac{1}{4}, \text{当 } x=0,1, \\ \frac{3}{4}, \text{当 } x=2,3.\end{cases} \tag{7.1.8}$$

也就是说，对于每个 x 值，选取 $\hat{p}(x)$使得

$$P(x,\hat{p}(x))\geqslant P(x,p(x)),$$

其中 $p(x)$是不同于 $\hat{p}(x)$ 的任一估计量. 这就是按极大似然法原理估计参数的基本思想.

更一般地，总体 ξ 为连续型随机变数，具有密度函数 $f(x;\theta)$，$\theta\in\Omega$，θ 为未知参数. 又设$(x_1,\cdots,x_n)$是子样 $(\xi_1,\cdots,\xi_n)$ 的一个观察值，那么子样$(\xi_1,\cdots,\xi_n)$落在点$(x_1,\cdots,x_n)$的邻域里的概率为$\prod_{i=1}^{n}f(x_i;\theta)\,dx_i$. 可见，$\theta$ 的取值不同，直接影响到$\prod_{i=1}^{n}f(x_i;\theta)\,dx_i$，因而它是 θ 的函数. 极大似然法原理就是选取使得子样落在观察值$(x_1,\cdots,x_n)$ 的邻域里的概率$\prod_{i=1}^{n}f(x_i;\theta)\,dx_i$ 达到最大的数值$\hat{\theta}(x_1,\cdots,x_n)$ 作为参数 θ 的估计值. 对于子样的每个观察值$(x_1,\cdots,x_n)$，都是选取使$\prod_{i=1}^{n}f(x_i;\theta)\,dx_i$ 达到最大的数值$\hat{\theta}(x_1,\cdots,x_n)$作为参数 θ 的估计值. 而 $\hat{\theta}(x_1,\cdots,x_n)$随子样观察值$(x_1,\cdots,x_n)$的不同而可能取不同的数值，因此 $\hat{\theta}$ 为子样的函数，可记作

$\hat{\theta}(\xi_1,\cdots,\xi_n)$. 当 ξ 是离散型随机变数时，用 ξ 的概率函数 $P(x;\theta)$ 代替 $f(x;\theta)$，可按上述原理同样进行讨论.

定义 7.1.3(极大似然法估计) 设总体 ξ 的密度函数为 $f(x;\theta_1,\cdots,\theta_l)$，其中 $\theta_1,\cdots,\theta_l$ 为未知参数，参数空间 Ω 是 l 维的. $\xi_1\cdots,\xi_n$ 为子样，它的联合密度函数为 $f(x_1,\cdots,x_n;\theta_1,\cdots,\theta_l)$. 称

$$L(\theta_1,\cdots,\theta_l)=\prod_{i=1}^{n}f(\xi_i;\theta_1,\cdots,\theta_l) \qquad (7.1.9)$$

为 $\theta_1,\cdots,\theta_l$ 的似然函数. 若有 $\hat{\theta}_1,\cdots,\hat{\theta}_l$ 使得下式成立:

$$L(\hat{\theta}_1,\cdots,\hat{\theta}_l)=\max_{(\theta_1,\cdots,\theta_l)\in\Omega}\{L(\theta_1,\cdots,\theta_l)\}, \quad (7.1.10)$$

则称 $\hat{\theta}_j=\hat{\theta}_j(\xi_1,\cdots,\xi_n)$ 为 θ_j 的极大似然法估计量 $(j=1,\cdots,l)$.

我们知道，

$$\ln L(\theta_1,\cdots,\theta_l)=\sum_{i=1}^{n}\ln f(\xi_i;\theta_1,\cdots,\theta_l),$$

由于 $\ln x$ 是 x 的单调上升函数，因而 $\ln L$ 与 L 有相同的极大值点. $\hat{\theta}_j$ 为极大似然估计量的必要条件为

$$\left.\frac{\partial \ln L(\theta_1,\cdots,\theta_l)}{\partial\theta_j}\right|_{\theta_j=\hat{\theta}_j}=0 \quad j=1,\cdots,l. \qquad (7.1.11)$$

称它为似然方程，由它解得 $\hat{\theta}_j=\hat{\theta}_j(\xi_1,\cdots,\xi_n)$，当满足(7.1.10)式时，则 $\hat{\theta}_j$ 为 θ_j 的极大似然法估计量)$j=1,\cdots,l$). ①

若 ξ 为离散型随机变数，概率函数为 $P(x;\theta_1,\cdots,\theta_l)$，则似然函数为

$$L(\theta_1,\cdots,\theta_l)=\prod_{i=1}^{n}P(\xi_i;\theta_1,\cdots,\theta_l). \qquad (7.1.12)$$

---

① 由数学分析求极值的原理知，(7.1.11)的解是否为极大值的点，还要验证二阶导数在该点是否小于零.

由似然方程组

$$\frac{\partial \ln L(\theta_1,\cdots,\theta_l)}{\partial \theta_j}\bigg|_{\theta_j=\hat{\theta}_j}=0 \quad j=1,\cdots,l$$

解得 $\theta_j=\theta_j(\xi_1,\cdots,\xi_n)$，若满足(7.1.10)式，则 θ_j 为 θ_j 的极大似然法估计量($j=1,\cdots,l$).

例 7.1.6 设总体 ξ 服从正态$N(a,\sigma)$，试求 a 及 σ 的极大似然法估计量.

解 我们知道，a 及 σ 的似然函数为

$$L(a,\sigma)=\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\xi_i-a)^2}{2\sigma^2}}\right]$$

$$=\frac{1}{(\sqrt{2\pi}\ \sigma)^n}e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(\xi_i-a)^2}.$$

似然方程为:

$$\begin{cases}\dfrac{\partial \ln L(a,\sigma)}{\partial a}=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(\xi_i-a)=0,\\[2ex]\dfrac{\partial \ln L(a,\sigma)}{\partial \sigma}=-\dfrac{n}{\sigma}+\dfrac{1}{\sigma^3}\sum\limits_{i=1}^{n}(\xi_i-a)^2=0.\end{cases}$$

解得:

$$\hat{a}=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\bar{\xi},$$

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=S^2.$$

则 $\bar{\xi}$ 及 S^2 分别为 a 及 σ^2 的极大似然法估计量.

例 7.1.7 设总体 ξ 服从具有参数 $\lambda>0$ 的泊松分布，概率函数为

$$P(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda},$$

$\lambda>0, x=0,1,2,\cdots$. 试求参数 λ 的极大似然估计量。

解 我们知道，参数 λ 的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\left[\frac{\lambda^{\xi_i}}{\xi_i!}e^{-\lambda}\right]=e^{-n\lambda}\frac{\lambda^{\sum\limits_{i=1}^{n}\xi_i}}{\prod\limits_{i=1}^{n}(\xi_i!)},$$

似然方程为

$$\frac{\partial\ln L(\lambda)}{\partial\lambda}=-n+\frac{1}{\lambda}\sum_{i=1}^{n}\xi_i=0.$$

解得

$$\lambda=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\bar{\xi},$$

即子样平均值 $\bar{\xi}$ 为参数 λ 的极大似然法估计量.

例 7.1.8 设总体 ξ 服从参数为 λ 的指数分布，密度函数为

$$f(x;\lambda)=\lambda e^{-\lambda x},$$

$x>0, \lambda>0$. 试求参数 λ 的极大似然法估计量。

解 我们知道，λ 的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}(\lambda e^{-\lambda\xi_i})=\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}\xi_i}$$

则

$$\ln L(\lambda)=n\ln\lambda-\lambda\sum_{i=1}^{n}\xi_i.$$

似然方程为

$$\frac{\partial\ln L(\lambda)}{\partial\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}\xi_i=0,$$

解得：

$$\lambda=\frac{1}{\bar{\xi}}.$$

例 7.1.9 设总体 ξ 在$[\theta_1,\theta_2]$上服从均匀分布，求 θ_1 及 θ_2 的极大似然估计量.

解 由 ξ 的密度函数

$$f(x;\theta_1,\theta_2)=\begin{cases}\dfrac{1}{\theta_2-\theta_1}, & x\in[\theta_1,\theta_2],\\ 0 & ,x\overline{\in}[\theta_1,\theta_2],\end{cases}$$

知 θ_1 及 θ_2 的似然函数为：

$$L(\xi_1,\cdots,\xi_n;\theta_1,\theta_2)=\begin{cases}\left(\dfrac{1}{\theta_2-\theta_1}\right)^n, & \theta_1\leqslant\xi_i\leqslant\theta_2,\ (i=1,2,\cdots,n),\\ 0 & ,\text{其它}.\end{cases}$$

似然方程为

$$\begin{cases}\dfrac{\partial\ln L}{\partial\theta_1}=\dfrac{n}{\theta_2-\theta_1}=0,\\ \dfrac{\partial\ln L}{\partial\theta_2}=\dfrac{-n}{\theta_2-\theta_1}=0.\end{cases}\tag{7.1.13}$$

我们看到，从(7.1.13)不可能解得 θ_1 及 θ_2 的极大似然估计量.

现在，我们从似然函数的定义出发来确定 θ_1 及 θ_2 的极大似然估计量. 我们知道，要使似然函数 $L(\xi_1,\cdots,\xi_n;\theta_1,\theta_2)$ 非零，必须要求 $\theta_1\leqslant\xi_1^*=\min\{\xi_1,\cdots,\xi_n\}$，$\xi_n^*=\max\{\xi_1,\cdots,\xi_n\}\leqslant\theta_2$，由于

$$L(\xi_1,\cdots,\xi_n;\theta_1,\theta_2)=\frac{1}{(\theta_2-\theta_1)^n}\leqslant\frac{1}{(\xi_n^*-\xi_1^*)^n},$$

今取

$$\hat{\theta}_1=\xi_1^*,\quad \hat{\theta}_2=\xi_n^*,\tag{7.1.14}$$

则有

$$L(\xi_1,\cdots,\xi_n;\theta_1,\theta_2)\leqslant L(\xi_1,\cdots,\xi_n;\hat{\theta}_1,\hat{\theta}_2).$$

因而 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 为 θ_1 及 θ_2 的极大似然估计量.

下面再举一个借助极大似然法的思想估计参数的例子.

例 7.1.10 钓鱼问题. 为了估计湖中的鱼数 N,同时自湖中钓出 r 条鱼做上记号后放回湖中,然后再自湖中同时钓出 s 条鱼,结果发现这 s 条中有 x 条标有记号. 这里N是未知数, r 及 s 是已知常数,试问应如何估计N的值?

解 第二次钓出的标有记号的鱼数 ξ 是随机变数, 且 ξ 服从超几何分布:

$$P\{\xi=x\}=\frac{C_r^x C_{N-r}^{s-x}}{C_N^s}, \tag{7.1.15}$$

其中 x 为整数, 且 $\max[0, s-(N-r)]\leqslant x\leqslant\min[r, s]$. 今用 $L(x;N)$表示(7.1.15)式右端,则取使 $L(x;N)$达到极大值的 $\hat{N}$ 作为N的估计量. 但是,直接对N求导数的方法相当困难,现用下述方法. 考虑比值

$$\begin{aligned}A(x;N)&=\frac{L(x;N)}{L(x;N-1)}=\frac{N-r}{N}\cdot\frac{N-s}{(N-r)-(s-x)}\\&=\frac{N^2-(r+s)N+rs}{N^2-(r+s)N+Nx}.\end{aligned} \tag{7.1.16}$$

从(7.1.16)式看到,当且仅当$N<\frac{rs}{x}$, $L(x;N)>L(x;N-1)$,当且仅当$N>\frac{rs}{x}$, $L(x;N)<L(x;N-1)$, 因此 $L(x;N)$在$\frac{rs}{x}$附近取极大值,于是N的估计量为:

$$\hat{N}=\left[\frac{rs}{x}\right]. \tag{7.1.17}$$

其右端的方括号表示取整数部分.

极大似然估计量有许多重要的性质. 因此, 当碰到似然方程难解时,为了获得参数的极大似然法估计量,设法求似然方程的近似解.

例 7.1.11 设总体 ξ 服从 Γ-分布 $\Gamma(\alpha,\beta)$, 求α及β的极

大似然估计量.

解 由(7.1.9)式知,参数α及β的似然函数为

$$L(\xi_1,\cdots,\xi_n;\alpha,\beta)=\frac{\beta^{-n(\alpha+1)}}{[\Gamma(\alpha+1)]^n}\left[\prod_{i=1}^{n}\xi_i\right]^{\alpha}e^{-\frac{1}{\beta}\sum\limits_{i=1}^{n}\xi_i}$$

则

$$\ln L=-n(\alpha+1)\ln\beta-n\ln\Gamma(\alpha+1)+\alpha\sum_{i=1}^{n}\ln\xi_i-\frac{1}{\beta}\sum_{i=1}^{n}\xi_i,$$

$$\frac{\partial\ln L}{\partial\alpha}=-n\ln\beta-\frac{n\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}+\sum_{i=1}^{n}\ln\xi_i,$$

$$\frac{\partial\ln L}{\partial\beta}=-\frac{n(\alpha+1)}{\beta}+\frac{n\bar{\xi}}{\beta^2}.$$

解得 $\hat{\beta}=\frac{\bar{\xi}}{\hat{\alpha}+1}$, 其中 $\hat{\alpha}$ 由下式确定:

$$\ln(\alpha+1)-\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)}=\ln\bar{\xi}-\frac{1}{n}\sum_{i=1}^{n}\ln\xi_i.\qquad(7.1.18)$$

D. G. Chapman 提出了解超越方程(7.1.18)的方法(1956), 并给出了函数 $\Gamma'(\alpha+1)/\Gamma(\alpha+1)$数值表供查用,参看[20].

记

$$Q(\alpha+1)=\ln(\alpha+1)-\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)},$$

$$G(\xi_1,\cdots,\xi_n)=\ln\bar{\xi}-\frac{1}{n}\sum_{i=1}^{n}\ln\xi_i,$$

用"试差法" 求α的估计$\hat{\alpha}$. 先给定α的初始值α_0, 若$Q(\alpha_0+1)=G$,则α_0为α的估计值. 如果$Q(\alpha_0+1)\neq G$,则选取α_1, 直至选出使等式成立或两端很接近的$\hat{\alpha}$值,作为α的估计值.

§7.2 无偏性与优效性

我们从上面已经看到，对于总体 ξ 的数学期望 $E(\xi)$，根据矩法可用子样的算术平均值 $\bar{\xi}$ 作为它的估计量. 但是，我们也可用另一种方法，例如用子样的加权平均值 $\sum_{i=1}^{n} c_i\xi_i$ 作为它的估计量，其中 $c_i \geqslant 0$，$\sum_{i=1}^{n} c_i = 1$. 甚至更简便地用容量为 1 的子样 ξ_1 作为 $E(\xi)$ 的估计. 问题在于哪个估计量为它的"最佳"估计，而"最佳"的准则又怎样确定？我们知道，未知参数 θ 的估计量 $T(\xi_1,\cdots,\xi_n)$ 为一不带有未知参数的统计量，是一随机变数. 而它的分布函数与总体 ξ 的分布函数有关，即与未知参数 θ 有关. 如果 $E_\theta(T)$ 等于或渐近于被估计的参数 θ，$D_\theta(T)$ 很小或渐近于零，那么估计量 $T(\xi_1,\cdots,\xi_n)$ 由子样观察值 $(x_1,\cdots,x_n)$ 而计算得到的数值，虽然不能都等于 θ，但其平均值等于或渐近于 θ，而且其离散度很小. 这时，用 $T(\xi_1,\cdots,\xi_n)$ 作为未知参数 θ 的估计，人们自然感到满意. 下面，我们从估计量 $T(\xi_1,\cdots,\xi_n)$ 的数学期望及方差这两个重要的数字特征出发，引进无偏性、优效性、相合性等概念，并引进充分性、完备性等概念，从不同角度来衡量估计量 $T(\xi_1,\cdots,\xi_n)$ 作为 θ 的"最佳"准则，即讨论估计量的优良性.

注意，并非未知参数 θ 的每个估计量都具有上述性质. 对于哥西分布而言，由于它的数学期望不存在，因而该参数估计量的无偏性及优效性将失去意义.

一、无偏性

定义 7.2.1(无偏估计量) 若参数 θ 的估计量 $T(\xi_1,\cdots,\xi_n)$ 对一切 n 及 $\theta\in\Omega$，有

$$E_\theta[T(\xi_1,\cdots,\xi_n)]=\theta, \tag{7.2.1}$$

则称 $T(\xi_1,\cdots,\xi_n)$ 为参数 θ 的无偏估计量. 记

$$E_\theta[T]-\theta=b_n, \tag{7.2.2}$$

称 b_n 为估计量 $T(\xi_1,\cdots,\xi_n)$ 的偏差. 若 $b_n\neq 0$, 则称 T 为 θ 的有偏估计量. 如果

$$\lim_{n\to\infty} b_n=0, \tag{7.2.3}$$

则称 T 为 θ 的渐近无偏估计量.

对于参数 θ 的任一实值函数 $g(\theta)$, 如果 $g(\theta)$ 的无偏估计量存在, 也就是说有估计量 T, 使得

$$E_\theta(T)=g(\theta), \tag{7.2.4}$$

则称 $g(\theta)$ 为可估计函数.

例 7.2.1 设总体 ξ 的 k 阶原点矩存在, 即 $\nu_k=E(\xi^k)$ 有限, 则子样的 k 阶原点矩为总体的 k 阶原点矩的无偏估计量.

证 设子样的 k 阶原点矩为

$$A_k=\frac{1}{n}\sum_{i=1}^{n}\xi_i^k,$$

则有

$$E(A_k)=\frac{1}{n}\sum_{i=1}^{n}E(\xi_i^k)=\nu_k. \tag{7.2.5}$$

特别地, 若 $E(\xi)=a$, 则 $A_1=\bar{\xi}$ 为 a 的无偏估计量.

例 7.2.2 设总体 ξ 的方差 $D(\xi)=\sigma^2$ 是有限的, 试求 σ^2 的无偏估计量.

解 子样的方差 S^2 是 σ^2 的矩法估计, 由于

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-(\bar{\xi})^2.$$

$$E(S^2)=E\left[\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-(\bar{\xi})^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}E(\xi_i^2)-E[(\bar{\xi})^2]=E(\xi^2)-[E(\bar{\xi})^2]$$

$$=\{D(\xi)+[E(\xi)]^2\}-\{D(\bar{\xi})+[E(\bar{\xi})]^2\}$$

$$=D(\xi)-D(\bar{\xi})=\sigma^2-D(\bar{\xi}).$$

其中

$$D(\bar{\xi})=D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(\xi)$$

$$=\frac{1}{n}D(\xi)=\frac{\sigma^2}{n},$$

于是

$$E(S^2)=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2. \tag{7.2.6}$$

因此，子样方差 S^2 不是总体的方差 σ^2 的无偏估计量，当然，S^2 是 σ^2 的渐近无偏估计量．现将有偏估计量 S^2 无偏化，记

$$S^{*2}=\frac{nS^2}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2, \tag{7.2.7}$$

则

$$E(S^{*2})=\sigma^2. \tag{7.2.8}$$

即 S^{*2} 为 σ^2 的无偏估计量．

由上述两例可见，不论总体 ξ 服从什么分布，只要 $E(\xi)=a$ 及 $D(\xi)=\sigma^2$ 是有限的话，那么

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i \quad S^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$$

分别为 a 及 σ^2 的无偏估计量．

注意，当 $E(\hat{\theta})=\theta$ 时，不一定有 $E[g(\hat{\theta})]=g(\theta)$，其中 $g(\theta)$ 为 θ 的实值函数．也就是说，当 $\hat{\theta}$ 为 θ 的无偏估计量时，$g(\hat{\theta})$ 不一定是 $g(\theta)$ 的无偏估计量．

例 7.2.3 设 $D(\xi)\neq0$，若 $E(\bar{\xi})=a$，则 $E(\bar{\xi}^2)=a^2$ 不成立

这是因为：

$$E(\bar{\xi}^2)=D(\bar{\xi})+[E(\bar{\xi})]^2=\frac{1}{n}D(\xi)+a^2,$$

所以

$$E(\bar{\xi}^2)\neq a^2.$$

定义 7.2.2（最小方差无偏估计量） 设 $T(\xi_1,\cdots,\xi_n)$ 为可估计函数 $g(\theta)$ 的无偏估计量，若对于 $g(\theta)$ 的任一无偏估计量 $T'(\xi_1,\cdots,\xi_n)$，有

$$D_\theta(T)\leqslant D_\theta(T') \qquad \text{对一切 } \theta\in\Omega,$$

则称 $T(\xi_1,\cdots,\xi_n)$ 为 $g(\theta)$ 的最小方差无偏估计量. 最小方差无偏估计量，简称为最优无偏估计量.

为便于下述两个定理的叙述，记

$$U=\{T: E_\theta(T)=\theta,\ D_\theta(T)<\infty,\ \text{对一切 } \theta\in\Omega\}, \quad (7.2.9)$$

即 U 为参数 θ 的方差有限的无偏估计量的集合.

又记

$$U_0=\{T_0: E_\theta(T_0)=0,\ D_\theta(T_0)<\infty,\ \text{对一切 } \theta\in\Omega\},$$

U_0 是参数 θ 的数学期望为零、方差有限的估计量的集合.

定理 7.2.1 设 U 是非空的集合，有一 $T\in U$，则 T 为 θ 的最优无偏估计量的充要条件为对每个 $T_0\in U_0$，有

$$E_\theta(T\cdot T_0)=0, \quad (7.2.10)$$

其中 $\theta\in\Omega$.

证 **必要性** 用反证法，设 $T\in U$ 为 θ 的最优无偏估计量，但条件(7.2.10)不满足，即存在 $T_0\in U_0$ 及 $\theta_0\in\Omega$，使得

$$E_{\theta_0}(T\cdot T_0)\neq0.$$

因为 $E_\theta(T_0)=0$，所以对一切 c 值，$T-cT_0\in U$. 于是

$$E_{\theta_0}(T-cT_0)^2=E_{\theta_0}(T^2)+c^2E_{\theta_0}(T_0^2)-2cE_{\theta_0}(T\cdot T_0).$$

由于 $E_{\theta_0}(T\cdot T_0)\neq 0$，一定能找到 c_0 值，例如

$$c_0=\frac{E_{\theta_0}(T\cdot T_0)}{E_{\theta_0}(T_0^2)}$$

使得

$$E_{\theta_0}(T-c_0T_0)^2<E_{\theta_0}(T^2),$$

所以

$$D_{\theta_0}(T-c_0T_0)<D_{\theta_0}(T).$$

这与 T 为最优无偏估计量的假定相矛盾，因此说明条件(7.2.10)是必要的.

充分性　设有某个 $T\in U$，使得(7.2.10)成立，我们来证 T 为 θ 的最优无偏估计量. 若(7.2.10)成立，则对任一估计量 $T'\in U$，显然 $T'-T\in U_0$，因而对一切 $\theta\in\Omega$，有

$$E_\theta[T(T'-T)]=0.$$

由许瓦兹不等式知，对一切 $\theta\in\Omega$，

$$E_\theta(T^2)=E_\theta(T\cdot T')\leqslant\sqrt{E_\theta(T^2)}\sqrt{E_\theta(T'^2)},$$

所以

$$E_\theta(T^2)\leqslant E_\theta(T'^2).$$

由于 $E_\theta(T)=E_\theta(T')=\theta$，因而对一切 $\theta\in\Omega$

$$D_\theta(T)\leqslant D_\theta(T').$$

可见 T 为 θ 的最优无偏估计量.

推论　设 T_1 和 T_2 分别是参数 θ 的可估计函数 $g_1(\theta)$ 和 $g_2(\theta)$ 的最优无偏估计量，则 $b_1T_1+b_2T_2$ 是 $b_1g_1(\theta)+b_2g_2(\theta)$ 的最优无偏估计量，其中 b_1 和 b_2 是常数.

证明留作练习(习题23).

定理 7.2.2　设 U 是(7.2.9)式所定义的非空集合，则对参数 θ 而言至多存在一个最优无偏估计.

证　假定 T 及 T' 是 θ 的两个最优无偏估计量，即对一切 $\theta\in$

Ω，有

$$E_\theta(T)=E_\theta(T')=\theta$$
$$D_\theta(T)=D_\theta(T').$$

因此对一切 $\theta\in\Omega$，有

$$E_\theta(T-T')=0.$$

即 $T-T'\in U_0$. 由定理 7.2.1 知，对一切 $\theta\in\Omega$，有：

$$E_\theta[T(T-T')]=0,$$
$$E_\theta[T'(T-T')]=0.$$

于是对一切 $\theta\in\Omega$，有：

$$E_\theta(T-T')^2=E_\theta[T(T-T')]-E_\theta[T'(T-T')]=0,$$

亦即对一切 $\theta\in\Omega$，有：

$$D_\theta(T-T')=0.$$

由车贝谢夫不等式知，对一切 $\theta\in\Omega$，有：

$$P_\theta\{T=T'\}=1.$$

这就说明参数 θ 的最优无偏估计量在概率为 1 的意义下是唯一的. ∎

例 7.2.4 设总体 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ^2 都是未知参数，试求 a 及 σ^2 的最优无偏估计量.

解 我们知道，若 $\xi_1,\cdots,\xi_n$ 为总体 ξ 的子样，则 $\bar{\xi}$，S^{*2} 分别为 a,σ^2 的无偏估计量，试问它们是否为最优无偏估计量？我们应用定理 7.2.1 看一看 $\bar{\xi}$、S^{*2} 是否满足(7.2.10)式的条件，即

$$E(\bar{\xi}\cdot T_0)=0,$$
$$E(S^{*2}\cdot T_0)=0$$

对每个 $T_0\in U_0$ 是否成立.

设 $T_0\in U_0$，则有 $E(T_0)=0$，即

$$\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x,\cdots,x_n)$$

$$\times \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0. \qquad (7.2.11)$$

将(7.2.11)对未知参数 a 求导数，则有

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x_1,\cdots,x_n)\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-a)\right]$$

$$\times \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0.$$

将非零系数消去，上式等价于

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}x_i\right]$$

$$\times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0.$$

于是有

$$E\left(T_0\cdot\sum_{i=1}^{n}\xi_i\right)=0,$$

即有

$$E(T_0\cdot\bar{\xi})=0.$$

由于 $\bar{\xi}$ 是 a 的无偏估计量. 应用定理 7.2.1 知，$\bar{\xi}$ 为参数 a 的最优无偏估计量.

将(7.2.11)式对参数 a 微分两次得到

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}x_i\right]^2$$

$$\times \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0,$$

将(7.2.11)对参数 σ^2 微分便得

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}(x_i-a)^2\right]$$

$$\times\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0$$

(其中积分号前的非零系数省略了). 由于

$$\sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-a)^2$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})^2+\frac{1}{n}\left(\sum_{i=1}^{n}x_i-na\right)^2,$$

因而可得

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T_0(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$\times\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]dx_1\cdots dx_n=0.$$

由于$\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=nS^2$是$(n-1)\sigma^2$的无偏估计量,应用定理7.2.1知$nS^2$是参数$(n-1)\sigma^2$的最优无偏估计量,再应用定理的推论知,$\frac{nS^2}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2=S^{*2}$是参数 σ^2 的最优无偏估计量.

我们说参数 θ 的估计量 $T(\xi_1,\cdots,\xi_n)$是子样 $\xi_1,\cdots,\xi_n$ 的线性函数,是指

$$T(\xi_1,\cdots,\xi_n)=\sum_{i=1}^{n}c_i\xi_i,$$

其中 $c_1,\cdots,c_n$ 为给定的常数.

定义 7.2.3(最小方差线性无偏估计) 可估计函数 $g(\theta)$ 的估计量 $T(\xi_1,\cdots,\xi_n)$ 称为它的最小方差线性无偏估计量,是指:如果

(1) $T(\xi_1,\cdots,\xi_n)$ 为子样的线性函数;

(2) $E_\theta[T(\xi_1,\cdots,\xi_n)]=g(\theta)$;

(3) 对于满足(1)及(2)的任一估计量 $T'(\xi_1,\cdots,\xi_n)$, 对一切 $\theta\in\Omega$ 有

$$D_\theta(T)\leqslant D_\theta(T').$$

定理 7.2.3 设 $T_1,\cdots,T_m$ 为可估计函数 $g(\theta)$ 的 m 个相互独立的线性无偏估计量, 而且它们有相同的方差 $D_\theta(T_j)=\sigma^2<\infty$, $j=1,\cdots,m$, 则统计量

$$T=\frac{1}{m}\sum_{j=1}^{m}T_j$$

是 $T_1,\cdots,T_m$ 的线性组合类中可估计函数 $g(\theta)$ 的最小方差线性无偏估计量, 且有 $D_\theta(T)=\frac{\sigma^2}{m}$.

证 令

$$T=\sum_{j=1}^{m}b_jT_j+b_0,$$

其中 b_0 及 $b_1,\cdots,b_m$ 为常数. 我们来考虑在什么情况下, 估计量 T 为 $g(\theta)$ 的最小方差线性无偏估计量. 首先要 T 为 $g(\theta)$ 的无偏估计量, 则由

$$E_\theta(T)=\sum_{j=1}^{m}b_jg(\theta)+b_0=g(\theta)\sum_{j=1}^{m}b_j+b_0$$

知 $b_0=0$, $\sum_{j=1}^{m}b_j=1$. 现进一步考虑估计量 T 的方差.

$$D_\theta(T)=\sum_{j=1}^{m}b_j^2E_\theta[T_j-E_\theta(T_j)]^2$$
$$=\sum_{j=1}^{m}b_j^2\sigma^2=\sigma^2\sum_{j=1}^{m}b_j^2.$$

欲要求 $D(T)$ 在 $\sum_{j=1}^{m}b_j=1$ 的约束条件下达到极小值，我们通过下述方法来讨论．记：

$$\varphi(c)=\sigma^2\left(\sum_{j=1}^{m}b_j^2\right)-c\left(\sum_{j=1}^{m}b_j-1\right),$$

c 为拉格朗日因子，由

$$\frac{\partial\varphi(c)}{\partial b_j}=0\quad(j=1,\cdots,m),$$

知

$$b_j=\frac{c}{2\sigma^2}\quad(j=1,\cdots,m).$$

又由 $\sum_{j=1}^{m}b_j=1$，因而 $c=\frac{2\sigma^2}{m}$，于是

$$b_j=\frac{1}{m}\quad(j=1,\cdots,m).$$

也就是说，当 $b_j=\frac{1}{m}$ 时，$D(T)$ 达到最小值，因而 $\bar{T}=\frac{1}{m}\sum_{j=1}^{m}T_j$ 是 T_1，$\cdots$，T_m 的线性组合类中 $g(\theta)$ 的最小方差线性无偏估计，且有 $D(\bar{T})=\sigma^2\sum_{j=1}^{m}\frac{1}{m^2}=\frac{\sigma^2}{m}$. ∎

推论 设总体 ξ 的数学期望 a 及方差 σ^2 都存在，$\xi_1,\cdots,\xi_n$ 为子样，则子样平均值 $\bar{\xi}$ 为 a 的最小方差线性无偏估计量.

证明留作练习.

最小方差线性无偏估计，简称作最优线性无偏估计．对于线性无偏估计类，有类似于定理 7.2.1 及定理 7.2.2 的结论．

二、优效性

我们知道，最优无偏估计量的方差总不会大于最优线性无偏估计量的方差．对于未知参数 θ，在它的许多不同的无偏估计量中，它们的方差可以小到怎样的程度（当然方差总是非负的）？另外，在参数 θ 的无偏估计类中，最优无偏估计量是否存在，怎样寻找最优无偏估计量？在线性无偏估计类中，又怎样寻找最优线性无偏估计量，这些问题在定理 7.2.1——定理 7.2.3 中已作了初步讨论．在这一段中，我们引进罗-克拉美不等式，指出无偏估计类中方差的下界，达到这个下界的无偏估计量称为优效估计量．

定理 7.2.4（罗-克拉美不等式）

设总体 ξ 为连续型随机变数，密度函数为 $f(x;\theta)$，θ 为未知参数，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为子样，$T(\xi_1,\cdots,\xi_n)$ 为可估计函数 $g(\theta)$ 的无偏估计量，如果

(1) $\dfrac{\partial f(x;\theta)}{\partial\theta}$ 存在，且有 $E_\theta\left[\dfrac{\partial}{\partial\theta}\ln f(\xi;\theta)\right]^2=I(\theta)>0$；

(2) $$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left[\prod_{i=1}^{n}f(x_i;\theta)\right]dx_1\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n}f(x_i;\theta)\right]dx_1\cdots dx_n;$$

(3) $\dfrac{\partial g(\theta)}{\partial\theta}$ 存在，且有

$$\frac{\partial g(\theta)}{\partial\theta}=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n}f(x_i;0)\right]dx_1\cdots dx_n$$

则有

$$D_\theta(T)\geqslant\frac{[g'(\theta)]^2}{nI(\theta)},\tag{7.2.12}$$

其中 $g'(\theta)=\frac{\partial g(\theta)}{\partial \theta}$. 特别当 $g(\theta)=\theta$ 时，上式可简化为

$$D_\theta(T)\geqslant\frac{1}{nI(\theta)}. \tag{7.2.13}$$

证明 若 $I(\theta)=\infty$ 或 $D_\theta(T)=\infty$，则(7.2.12) 式显然成立，因而可假定 $I(\theta)<\infty, D_\theta(T)<\infty$.

首先，我们知道

$$\int_{-\infty}^{\infty}f(x_i;\theta)\,dx_i=1 \quad (i=1,2,\cdots,n);$$

$$g(\theta)=E(T)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\prod_{i=1}^{n}[f(x_i;\theta)\,dx_i].$$

将上述两式分别对 θ 求微商，则有

$$0=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x_i;\theta)\,dx_i=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}[\ln f(x_i;\theta)]f(x_i;\theta)\,dx_i$$

$$=E\left[\frac{\partial}{\partial\theta}\ln f(\xi_i;\theta)\right] \quad (i=1,\cdots,n);$$

$$g'(\theta)=\frac{\partial}{\partial\theta}\left\{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\prod_{i=1}^{n}[f(x_i;\theta)\,dx_i]\right\}$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}T(x_1,\cdots,x_n)\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\right]$$

$$\times\prod_{i=1}^{n}[f(x_i;\theta)\,dx_i]$$

$$=E[T\cdot Z]=E(T)E(Z)+\mathrm{cov}(T,Z)$$

$$=E(T)E(Z)+\rho\sqrt{D(T)}\sqrt{D(Z)}.$$

其中

$$Z=\sum_{i=1}^{n}\frac{\partial}{\partial\theta}[\ln f(\xi_i;\theta)], \quad T=T(\xi_1,\cdots,\xi_n),$$

$$\rho=\frac{\mathrm{cov}(T,Z)}{\sqrt{D(T)}\sqrt{D(Z)}}.$$

由于

$$E(Z)=\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta}\ln f(\xi_i;\theta)\right]=0,$$

$$D(Z)=E(Z^2)=\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta}\ln f(\xi_i;\theta)\right]^2$$

$$=nE\left[\frac{\partial}{\partial\theta}\ln f(\xi_i;\theta)\right]^2,$$

于是

$$\rho^2=\frac{[g'(\theta)]^2}{D(T)\cdot D(Z)}.$$

因为 $\rho^2\leqslant 1$，所以

$$D(T)\geqslant\frac{[g'(\theta)]^2}{D(Z)}$$

$$=\frac{[g'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right]^2}. \blacksquare$$

若总体 ξ 为离散型，其概率函数 $P(x;\theta)$ 满足定理 7.2.4 中相应的(1)、(2)及(3)三个条件，则有：

$$D(T)\geqslant\frac{[g'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\ln P(\xi;\theta)\right]^2}.$$

定义 7.2.4(正规分布)　若总体 ξ 的密度函数（或概率函数）满足定理 7.2.4 中的(1)及(2)两个条件，则称 ξ 的分布函数为正规分布，满足条件(3)的估计量称为正规估计量.

定义 7.2.5(优效估计)　如果参数 θ 的函数 $g(\theta)$ 的某个正规无偏估计量 $U(\xi_1,\cdots,\xi_n)$ 的方差达到罗-克拉美不等式的下界，则称这个无偏估计量 $U(\xi_1,\cdots,\xi_n)$ 为 $g(\theta)$ 的优效估计量.

就是说，参数 $g(\theta)$ 的正规无偏估计量 $U(\xi_1,\cdots,\xi_n)$ 是优效估

计量，当且仅当$U(\xi_1,\cdots,\xi_n)$的方差为：

$$D(U)=\frac{[g'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right]^2}. \tag{7.2.14}$$

这是连续型随机变数ξ的情形；当ξ为离散型时，则为

$$D(U)=\frac{[g'(\theta)]^2}{nE\left[\frac{\partial}{\partial\theta}\ln P(\xi;\theta)\right]^2}. \tag{7.2.15}$$

定义 7.2.6（有效率） 设$T(\xi_1,\cdots,\xi_n)$为参数$g(\theta)$的任一正规无偏估计量，$U(\xi_1,\cdots,\xi_n)$为$g(\theta)$的优效估计量，则称

$$e_n=\frac{D(U)}{D(T)} \tag{7.2.16}$$

为正规无偏估计量$T(\xi_1,\cdots,\xi_n)$的有效率.

易见，有效率e_n满足关系式：$0\leqslant e_n\leqslant 1$. 如果正规无偏估计量$T$的有效率$e_n$满足：

$$\lim_{n\to\infty}e_n=1, \tag{7.2.17}$$

则称$T(\xi_1,\cdots,\xi_n)$为参数θ的渐近优效估计量.

可以证明，(7.2.12)式等价于下式：

$$D(T)\geqslant\frac{-[g'(\theta)]^2}{nE\left[\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)\right]}. \tag{7.2.18}$$

（见习题22）.

例 7.2.5 设总体ξ服从$\Gamma(\alpha,\beta)$，其中α为已知常数，试求参数β的优效估计.

解 由例7.1.3及例7.1.11知，$\hat{\beta}=\frac{\bar{\xi}}{\alpha+1}$. 我们来证$\hat{\beta}=\frac{\bar{\xi}}{\alpha+1}$为$\beta$的优效估计量. 由于

$$E(\xi)=\beta(\alpha+1),\quad D(\xi)=\beta^2(\alpha+1),$$

所以

$$E(\bar{\xi})=\beta(\alpha+1),\quad D(\bar{\xi})=\frac{\beta^2(\alpha+1)}{n},$$

则有

$$E(\beta)=\beta,\quad D(\beta)=\frac{\beta^2}{n(\alpha+1)}.$$

可以验证，估计量 β 是正规的. 现计算

$$n\int_0^{\infty}\left[\frac{\partial}{\partial\beta}\ln f(x,\beta)\right]^2 f(x,\beta)\,dx$$

$$=n\int_0^{\infty}\left[\frac{x}{\beta^2}-\frac{\alpha+1}{\beta}\right]^2 f(x,\beta)\,dx$$

$$=n\beta^{-4}E[\xi-E(\xi)]^2=n\beta^{-4}D(\xi)=n\beta^{-2}(\alpha+1).$$

今取 $g(\beta)=\beta$. 由 (7.2.14) 式知，β 的优效估计的方差应该等于

$$\frac{1}{nE\left[\frac{\partial}{\partial\beta}\ln f(x,\beta)\right]^2}=\frac{\beta^2}{n(\alpha+1)}.$$

由此可见，$\beta=\frac{\bar{\xi}}{\alpha+1}$ 为 β 的优效估计.

例 7.2.8 设总体 ξ 服从正态 $N(a,\sigma^2)$，a 及 σ 都是未知参数，$\xi_1,\cdots,\xi_n$ 为子样，则 S^{*2} 为 σ^2 的渐近优效估计量，其中

$$S^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2.$$

证 由定理 6.3.1 知，$\frac{(n-1)S^{*2}}{\sigma^2}\sim\chi^2_{(n-1)}$，于是 $E\left[\frac{(n-1)S^{*2}}{\sigma^2}\right]=n-1$，$D\left[\frac{(n-1)S^{*2}}{\sigma^2}\right]=2(n-1)$，因而

$$E(S^{*2})=\sigma^2,\quad D(S^{*2})=\frac{2\sigma^4}{n-1}.$$

由不等式(7.2.18)的右端，计算优效估计量的方差. 在本例

中，$\theta=\sigma^2$，考虑 $g(\theta)=\theta$ 的情形，因而(7.2.18)式右端的分子为1，现在来求分母.

$$\ln f(x;a,\sigma^2)=-\frac{(x-a)^2}{2\sigma^2}-\ln\sqrt{2\pi\sigma^2},$$

$$\frac{\partial}{\partial\sigma^2}\ln f(x;a,\sigma^2)=\frac{(x-a)^2}{2\sigma^4}-\frac{1}{2\sigma^2},$$

$$\frac{\partial}{\partial\sigma^2}\left[\frac{\partial}{\partial\sigma^2}\ln f(x;a,\sigma^2)\right]=-\frac{(x-a)^2}{\sigma^6}+\frac{1}{2\sigma^4},$$

所以

$$-E\left\{\frac{\partial}{\partial\sigma^2}\left[\frac{\partial}{\partial\sigma^2}\ln f(\xi;a,\sigma^2)\right]\right\}=\frac{1}{2\sigma^4}.$$

由此可见，参数 σ^2 的优效估计量的方差为$\frac{2\sigma^4}{n}$，那么 S^{*2} 的有效率 $e_n=\frac{n-1}{n}$，可见 S^{*2} 为 σ^2 的渐近优效估计量.

三、相合性

在无偏估计类中，我们以估计量的方差大小作为衡量估计量为"最优"的准则，作了较为充分的讨论. 但是，无偏估计类中方差为最小或较小的估计量，不一定比某个有偏差的估计量的方差来得小. 无偏与有偏是反映估计量 $\hat{\theta}(\xi_1,\cdots,\xi_n)$ 的数学期望是否等于被估计的真参数值 θ；方差 $D(\hat{\theta})$ 的大小反映 $\hat{\theta}$ 的观察值以真参数值 θ 为中心的离散程度. 一个估计量，它依赖于子样观察值而求得的数值，即使其平均值等于 θ，但离散程度很大，那么这个估计量用来估计 θ 时，仍然不大理想. 因此，人们希望在偏差性（有偏或无偏）与离散性（方差的大小）两者兼顾的原则下来建立估计量为"最优"的准则，为此引进相合性的概念.

例 7.2.7 设总体 ξ 在 $(0,\theta)$ 上服从均匀分布，密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta,\\ 0, & \text{其它}.\end{cases}$$

我们可以求出参数 θ 的矩法估计量 $\hat{\theta}_1=2\bar{\xi}$，且有

$$E(\hat{\theta}_1)=E(2\bar{\xi})=\theta,$$

$$D(\hat{\theta}_1)=D(2\bar{\xi})=4D(\bar{\xi})$$

$$=\frac{4}{n}D\ (\xi)=\frac{\theta^2}{3n}.$$

我们又可求出参数 θ 的极大似然法估计量 $\hat{\theta}_2=\max\limits_{1\leq i\leq n}(\xi_1,\cdots,\ \xi_n)$，且有

$$E(\hat{\theta}_2)=\frac{n\theta}{n+1}, \tag{7.2.19}$$

$$D(\hat{\theta}_2)=\frac{n\theta^2}{(n+1)^2(n+2)}. \tag{7.2.20}$$

可见 $\hat{\theta}_2(\xi_1,\cdots,\xi_n)$是 θ 的渐近无偏估计量. (7.2.19)及(7.2.20)两式的证明留作练习(见习题 24).

从上述等式看到，$\hat{\theta}_1$ 是无偏估计量，$\hat{\theta}_2$ 是有偏估计量，但 $\hat{\theta}_2$的方差比 $\hat{\theta}_1$ 来得小，这是因为在 $n\geqslant1$ 时，有

$$\frac{n}{(n+1)^2(n+2)}<\frac{1}{3n},$$

特别当子样容量 n 很大时更为显著，因为这时

$$\lim_{n\to\infty}\frac{D(\hat{\theta}_2)}{D(\hat{\theta}_1)}=\lim_{n\to\infty}\frac{3n^2}{(n+1)^2(n+2)}=0.$$

例 7.2.8 设总体 ξ 服从正态 $N(a,\sigma)$，$\xi_1,\cdots,\xi_n$ 为子样，a及 σ 都为未知参数，我们用统计量

$$S_k^2=\frac{1}{k}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2 \tag{7.2.21}$$

作为 σ^2 的估计量，其中 $\bar{\xi}$ 为子样平均值. 由定理 6.3.1 知

$$\frac{kS_k^2}{\sigma^2}\sim\chi^2{}_{(n-1)},$$

所以

$$E\left(\frac{kS_k^2}{\sigma^2}\right)=n-1,$$

$$D\left(\frac{kS_k^2}{\sigma^2}\right)=2(n-1).$$

因而

$$E(S_k^2)=\frac{n-1}{k}\sigma^2, \tag{7.2.22}$$

$$D(S_k^2)=\frac{2(n-1)}{k^2}\sigma^4. \tag{7.2.23}$$

特别地当 $k=n-1$ 时，则有

$$E(S_{n-1}^2)=\sigma^2,$$

$$D(S_{n-1}^2)=\frac{2\sigma^4}{n-1}.$$

从(7.2.22) 及 (7.2.23) 两式看到，估计量 S_k^2 的偏差 $b_k=\left[\frac{n-1}{k}-1\right]\sigma^2$，当 k 取值自 $n-1$ 增大时，偏差 b_k 的绝对值增大，而方差却减小. 例如

$$E(S_{n+1}^2)=\frac{n-1}{n+1}\sigma^2,$$

$$D(S_{n+1}^2)=\frac{2(n-1)}{(n+1)^2}\sigma^4,$$

$D(S_{n+1}^2)<D(S_{n-1}^2)$，但 S_{n+1}^2 是 σ^2 的有偏估计量. 试问：怎样衡量 S_{n-1}^2 与 S_{n+1}^2 哪个为 σ^2 的较优估计量？

对于(7.2.21)式所定义的 σ^2 的估计量 S_k^2，我们可计算估计量 S_k^2 与被估计参数 σ^2 之间的均方误差为：

$$\begin{aligned}E[S_k^2-\sigma^2]^2&=E[S_k^2-E(S_k^2)+E(S_k^2)-\sigma^2]^2\\&=D(S_k^2)+b_k^2,\end{aligned} \tag{7.2.24}$$

其中

$$E\{[S_k^2-E(S_k^2)][E(S_k^2)-\sigma^2]\}$$

$$=[E(S_k^2)-\sigma^2]E[S_k^2-E(S_k^2)]=0.$$

于是

$$E[S_k^2-\sigma^2]^2=\frac{2(n-1)}{k^2}\sigma^4+\left(\frac{n-1}{k}-1\right)^2\sigma^4. \qquad (7.2.25)$$

当 $k=n-1$ 时，有

$$E[S_{n-1}^2-\sigma^2]^2=\frac{2}{n-1}\sigma^4,$$

当 $k=n+1$ 时，有

$$E[S_{n+1}^2-\sigma^2]^2=\frac{2}{n+1}\sigma^4,$$

于是

$$E[S_{n+1}^2-\sigma^2]^2<E[S_{n-1}^2-\sigma^2]^2.$$

(7.2.24)式的均方误差，它由估计量的方差及偏差平方两项组成，当左端达到最小时，其方差及偏差的绝对值都有可能很小，特别是在左端趋于零时，右端的两个非负项都分别趋于零. 在均方误差的数学期望为最小的原则下，参数 σ^2 的估计量取 S_{n+1}^2 较 S_{n-1}^2 为优，而这个原则是在兼顾了方差及偏差绝对值都相对地小的情况下确定的.

定义 7.2.7 相合性估计. 设统计量 $T(\xi_1,\cdots,\xi_n)$ 为待估函数 $g(\theta)$ 的估计量，若对一切 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|T(\xi_1,\cdots,\xi_n)-g(\theta)|\leqslant\varepsilon\}=1,$$

则称 $T(\xi_1,\cdots,\xi_n)$ 为 $g(\theta)$ 的弱相合估计. 如果下式成立：

$$P\{\lim_{n\to\infty}T(\xi_1,\cdots,\xi_n)=g(\theta)\}=1,$$

则称 $T(\xi_1,\cdots,\xi_n)$ 为 $g(\theta)$ 的强相合估计.

我们知道，对于 $g(\theta)$ 的任何估计量 $T(\xi_1,\cdots,\xi_n)$，下式都成立：

$$E[T(\xi_1,\cdots,\xi_n)-g(\theta)]^2=D(T)+b_n^2, \qquad (7.2.26)$$

由于对任给 $\varepsilon>0$，下式成立

$$P\{|\hat{\theta}-\theta|>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E(\hat{\theta}-\theta)^2,$$

所以当 $n\to\infty$ 时，若方差 $D(T)$ 及偏差 b_n 都趋于零，则有

$$\lim_{n\to\infty}P\{|T(\xi_1,\cdots,\xi_n)-g(\theta)|>\varepsilon\}=0.$$

因此，$T(\xi_1,\cdots,\xi_n)$ 为 $g(\theta)$ 的弱相合估计.

如果估计量 $\hat{\theta}=T(\xi_1,\cdots,\xi_n)$ 为参数 θ 的相合性估计量，则 $\hat{\theta}$ 依概率收敛于 θ，即对任给 $\varepsilon>0$，有：

$$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|>\varepsilon\}=0. \qquad (7.2.27)$$

或等价地有

$$\lim_{n\to\infty}P\{|\hat{\theta}-\theta|\leqslant\varepsilon\}=1.$$

在有些教材中，常称 $\hat{\theta}$ 为参数 θ 的一致估计量.

相合性或一致性的概念，是在极限的意义下引进的，适用于大样本情形

*§7.3 充分性与完备性

在前两节中引进了求参数点估计的矩法与极大似然法，讨论了点估计量的无偏性与优效性，这是点估计的最基本知识．从这两节的内容讨论中，人们会发现还有一些问题，有必要给予进一步的探讨．如对于所获得的容量为 n 的子样 $\xi_1,\cdots,\xi_n$，常用子样的一个函数作为参数的点估计，试问这样建立点估计有没有充分利用子样所获得的信息？也就是说，用矩法与极大似然法确定参数的点估计，在哪些情况下充分利用了子样获得的信息？又如，怎样寻找参数的最小方差无偏估计及优效估计？在这节里，将引进充分性与完备性的概念，从而完整地解决寻找最小方差无偏估计及优效估计的问题.

充分性与完备性是数理统计学中很重要的概念，在估计理论与假设检验中都很有用.

一、充分性

定义 7.3.1 （充分统计量）设总体 ξ 的分布函数为 $F(x,\theta)$，θ 为未知参数，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为子样，$T(\xi_1,\cdots,\xi_n)$ 为不带有未知参数的统计量，若在给定统计量 $T(x_1,\cdots,x_n)=t$ 时，$(\xi_1,\cdots,\xi_n)$ 的条件分布 $F(x_1,\cdots,x_n|t)$ 与 θ 无关，则称 $T=T(\xi_1,\cdots,\xi_n)$ 为 θ 的充分统计量.

例 7.3.1 设总体 ξ 服从两点分布，即有

$$P\{\xi=x\}=\begin{cases}p, & 当\ x=1,\\ 1-p, & 当\ x=0,\end{cases}$$

其中 $0<p<1$. $\xi_1,\cdots,\xi_n$ 为子样，则子样平均值 $\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 为参数 p 的充分统计量.

证 由于

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\frac{k}{n}$$

其中 k 为子样 $\xi_1,\cdots,\xi_n$ 中取值为 1 的个数，则子样关于 $\bar{\xi}$ 的条件分布为

$$P\left\{\xi_1=e_1,\cdots,\xi_n=e_n\,|\,\bar{\xi}=\frac{k}{n};p\right\}$$

$$=\frac{P\left\{\xi_1=e_1,\cdots,\xi_n=e_n;\ \bar{\xi}=\frac{k}{n};p\right\}}{P\left\{\bar{\xi}=\frac{k}{n};p\right\}}$$

$$=\frac{P\left\{\xi_1=e_1,\cdots,\xi_n=e_n;\sum_{i=1}^{n}\xi_i=k;p\right\}}{P\left\{\sum_{i=1}^{n}\xi_i=k;p\right\}}$$

$$=\frac{p^k(1-p)^{n-k}}{C_n^k p^k(1-p)^{n-k}}=\frac{1}{C_n^k},$$

其中 $x_i=1$ 或 0，$i=1$，…，n. 于是在给定统计量 $\bar{\xi}$ 的取值为 $\frac{k}{n}$ 时，子样的条件分布与参数 p 无关，因而 $\bar{\xi}$ 为 p 的充分统计量.

充分统计量的直观背景是很有意思的. 我们知道，容量为 n 的子样 $\xi_1,\cdots,\xi_n$ 的观察值 $x_1,\cdots,x_n$ 中每个 x_i 都反映了总体 ξ 的特征，而在用子样平均 $\bar{\xi}$ 估计 $E(\xi)$ 时，只用 $\bar{\xi}$ 的观察值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 作为它的估计值. 就是说，对 n 个数据 $x_1,\cdots,x_n$ 进行了"压缩"和"精简"，试问这样做是否合理？亦即 $\bar{\xi}$ 是否已经包含了子样关于 $E(\xi)$ 的全部信息？或者说 $\bar{\xi}$ 作为 $E(\xi)$ 的估计是否充分利用了子样关于 $E(\xi)$ 的全部信息？从几何学上来看，子样的观察值 $(x_1,\cdots,x_n)$ 是 n 维的子样空间 $\mathscr{X}$ 中的一点，而当已知 $\bar{\xi}=\bar{x}$ 时，子样 $(\xi_1,\cdots,\xi_n)$ 是 $n-1$ 维空间的点了，就是说在已知 $\bar{\xi}=\bar{x}$ 时，进一步知道 $(\xi_1,\cdots,\xi_n)$ 的各自取何值，对于被估计的 $E(\xi)$ 来说，能否获得更多的信息？为说明简便，我们就 $n=2$ 的特殊情形，举例说明.

例 7.3.2 设总体 ξ 服从正态 $N(a,1)$，要求估计 a. ξ_1,ξ_2 为总体 ξ 的子样，用 $\bar{\xi}=\frac{1}{2}(\xi_1+\xi_2)$ 估计 a. 此时，子样空间 $\mathscr{X}$ 可用平面来表示，那么在给定 $\bar{\xi}$ 的取值为 y 时，$\frac{1}{2}(x_1+x_2)=y$，表示该平面 (x_1,x_2) 上的一条直线. 在这平面上，(ξ_1,ξ_2) 服从二维正态分布，其密度函数为

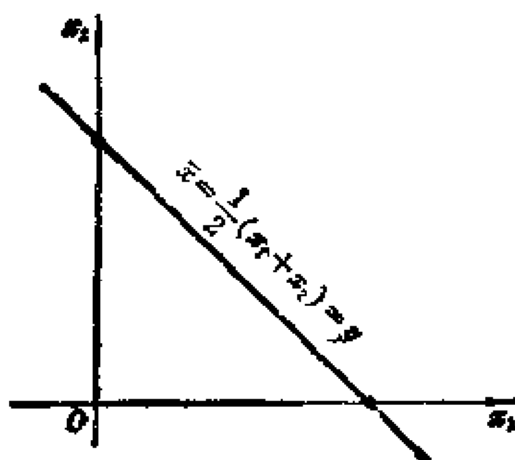


图 7.3.1

$$f(x_1,x_2;a)=\frac{1}{2\pi}\exp\left[-\frac{1}{2}\sum_{i=1}^{2}(x_i-a)^2\right],$$

子样平均值 $\bar{\xi}$ 服从正态 $N\left(a,\frac{1}{\sqrt{2}}\right)$，其密度函数为

$$f(\bar{x};a)=\frac{1}{\sqrt{\pi}}\exp[-(\bar{x}-a)^2].$$

在直线 $y=\frac{1}{2}(x_1+x_2)$ 上，(ξ_1,ξ_2) 的分布可以用条件分布密度函数

$$f(x_1,x_2|\bar{x}=y)=\begin{cases}\frac{f(x_1,x_2)}{f(y)}, & \text{当 } \bar{x}=y,\\ 0, & \text{当 } \bar{x}\neq y,\end{cases}$$

来描述，其中 $\bar{x}$ 为 $\bar{\xi}$ 的观察值，而 y 为 $\bar{\xi}$ 的给定值. 所以

$$f(x_1,x_2|\bar{x}=y)$$

$$=\begin{cases}\frac{1}{2\sqrt{\pi}}\exp\left[-\frac{1}{2}\sum_{i=1}^{2}(x_i-y)^2\right], & \text{当 } \bar{x}=y,\\ 0 & ,\text{当 } \bar{x}\neq y,\end{cases}$$

其中 $\bar{x}=\frac{1}{2}(x_1+x_2)$.

从上面条件密度函数的表示式中看到，它与参数 a 无关，这表明在已知 $\bar{x}=y$ 时，进一步知道子样观察值 (x_1,x_2) 在这条直线上的具体位置(即 x_1 及 x_2 各取何值使得 $\bar{x}=y$)，丝毫没有增加关于参数 a 的任何新的信息. 换句话说，$\bar{x}$ 已充分地从子样 ξ_1,ξ_2 中提取了所包含的关于参数 a 的全部信息，因而称 $\bar{\xi}=\frac{1}{2}\sum_{i=1}^{2}\xi_i$ 为参数 a 的充分统计量.

若用参数 θ 的充分统计量 $T(\xi_1,\cdots,\xi_n)$ 作为 θ 的点估计，则称 $T(\xi_1,\cdots,\xi_n)$ 为参数 θ 的充分估计量.

定理 7.3.1(费歇-奈曼因子分解定理) 若总体 ξ 为连续型随

机变数，密度函数为 $f(x,\theta)$，$\xi_1,\cdots,\xi_n$ 为子样，则统计量 $T(\xi_1,\cdots,\xi_n)$ 为参数 θ 的充分估计量的充要条件是：子样的联合密度函数可表示为

$$\prod_{i=1}^{n} f(x_i,\theta)=f_T(t,\theta)h(x_1,\cdots,x_n), \qquad (7.3.1)$$

其中 $f_T(t,\theta)$ 是仅通过 $T(x_1,\cdots,x_n)=t$ 而依赖于子样且与 θ 有关的非负函数，$h(x_1,\cdots,x_n)$ 是子样 $\xi_1,\cdots,\xi_n$ 的非负函数但与参数 θ 无关. 特别地，$f_T(t,\theta)$ 可为统计量 T 的密度函数.

若总体 ξ 为离散型随机变数，则密度函数代之以概率函数 $P(x,\theta)$，有类似于(7.3.1)的分解式.

这个定理的证明在此从略。读者可参看[20]的定理 1.5.1 及定理 1.5.2.

如果 θ 是参数向量，如正态 $N(a,\sigma)$ 中的 a 及 σ^2 都是未知参数，则记 $\theta=(a,\sigma^2)$；统计量 $T(\xi_1,\cdots,\xi_n)$ 是随机向量，而且定理7.3.1 的条件相应地成立，则称 T 为关于 θ 的联合充分统计量.

例 7.3.3 设总体 ξ 服从如下均匀分布

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta}, & 0<x<\theta,\\ 0, & 其它,\end{cases}$$

$\xi_1,\cdots,\xi_n$ 为子样，求参数 θ 的充分估计量.

解 子样的联合密度函数为

$$\prod_{i=1}^{n} f(x_i,\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 当 0\leqslant \max\limits_{1\leqslant i\leqslant n} x_i\leqslant\theta\\ 0, & 其它\end{cases}$$

$$=\frac{1}{\theta^n}\chi[0\leqslant \max_{1\leqslant i\leqslant n} x_i\leqslant\theta],$$

其中 χ 表示集合 $[0\leqslant \max\limits_{1\leqslant i\leqslant n} x_i\leqslant\theta]$ 的示性函数. 由定理 7.3.1 知，

统计量 $T(\xi_1,\cdots,\xi_n)=\max\limits_{1\leqslant i\leqslant n}\xi_i=\xi_n^*$ 为 θ 的充分估计量.

例 7.3.4 设总体 ξ 服从泊松分布 $P(x,\lambda)$，$\xi_1,\cdots,\xi_n$ 为子样，求参数 λ 的充分估计量.

解 子样的联合概率函数为

$$P(x_1,\cdots,x_n;\lambda)=\frac{e^{-n\lambda}}{x_1!\cdots x_n!}\lambda^{\sum\limits_{i=1}^{n}x_i}.$$

取

$$T(\xi_1,\cdots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i=\bar{\xi};$$

$$h(x_1,\cdots,x_n)=\frac{1}{x_1!\cdots x_n!};$$

$$f_T(t,\lambda)=\lambda^{nt}e^{-n\lambda}.$$

则有

$$P(x_1,\cdots,x_n;\lambda)=f_T(t,\lambda)h(x_1,\cdots,x_n).$$

因而统计量 $T=\bar{\xi}$ 为参数 λ 的充分估计量.

例 7.3.5 设总体 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是未知参数，求 $\theta=(a,\sigma^2)$ 的充分估计量.

解 子样的联合密度函数为

$$\prod_{i=1}^{n}f(x_i,\theta)=\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-a)^2\right]$$

$$=\frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left\{-\frac{n}{2\sigma^2}[s^2+n(\bar{x}-a)^2]\right\}$$

记二维统计量 $T(\xi_1,\cdots,\xi_n)=(\bar{\xi},s^2)$，由定理 7.3.1 知，$(\bar{\xi},s^2)$ 是参数向量 $\theta=(a,\sigma^2)$ 的充分估计量.

定理 7.3.2 参数 θ 的正规无偏估计量 $T(\xi_1,\cdots,\xi_n)$ 是优效估计量的充要条件为

(1) $T(\xi_1,\cdots,\xi_n)$为 θ 的充分估计量.

(2) 统计量 T 的密度函数(或联合概率函数,视总体 ξ 为连续型或离散型随机变数而定)概率为 1 地满足下述方程:

$$\frac{\partial \ln f_T(t,\theta)}{\partial\theta}=C(\theta)[T(x_1,\cdots,x_n)-\theta], \qquad (7.3.2)$$

其中 $C(\theta)$ 与子样无关,$T(x_1,\cdots,x_n)$ 与 θ 无关.

证 **必要性** 假定 $T(\xi_1,\cdots,\xi_n)$是 θ 的优效估计量,由 (7.2.13)式有:

$$D_\theta(T)=\frac{1}{nI(\theta)}.$$

我们来证条件(1)和(2)是必要的. 在不等式(7.2.13)的证明过程中看到,等号成立的条件是 T 与 Z 的相关系数 $|\rho|=1$, 即 T 与 Z 之间的线性关系式

$$Z=C(\theta)[T-\theta]$$

几乎处处成立,其中 $C(\theta)$ 只是与参数 θ 有关的数,不是统计量,也就是

$$Z=\frac{\partial \ln L(\xi_1,\cdots,\xi_n;\theta)}{\partial\theta}=C(\theta)[T-\theta].$$

将上式在(θ_0,θ)上关于 θ 积分,θ_0 为参数空间 Ω 中的任一固定值,得到

$$L(\xi_1,\cdots,\xi_n;\theta)=\exp\{T\cdot\varphi_1(\theta)-\varphi_2(\theta)\},$$

其中

$$\varphi_1(\theta)=\int_{\theta_0}^{\theta}C(\theta)d\theta, \quad \varphi_2(\theta)=\int_{\theta_0}^{\theta}\theta C(\theta)d\theta.$$

由定理 7.3.1 知,$T(\xi_1,\cdots,\xi_n)$为 θ 的充分估计量. 因而

$$\frac{\partial \ln f_T(t,\theta)}{\partial\theta}=C(\theta)[T(x_1,\cdots,x_n)-\theta].$$

充分性 如果条件(1)及(2)满足,则由定理 7.3.1 知

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\frac{\partial}{\partial\theta}\left[\ln\prod_{i=1}^{n}f(x_i,\theta)\right]\right\}^2\prod_{i=1}^{n}[f(x_i,\theta)\,dx_i]$$

$$=C^2(\theta)\int_{-\infty}^{\infty}[t-\theta]^2 f_T(t,\theta)\,dt=C^2(\theta)D_\theta(T),$$

$$1=\int_{-\infty}^{\infty}t\frac{\partial}{\partial\theta}[f_T(t,\theta)]\,dt=\int_{-\infty}^{\infty}[t-\theta]\frac{\partial}{\partial\theta}[f_T(t,\theta)]\,dt$$

$$=\int_{-\infty}^{\infty}[t-\theta]\frac{\partial}{\partial\theta}[\ln f_T(t,\theta)]f_T(t,\theta)\,dt$$

$$=C(\theta)D_\theta(T),$$

因此,

$$nI(\theta)=\frac{1}{D_\theta(T)}.\blacksquare$$

定理 7.3.3 设总体 ξ 的分布函数 $F(x;\theta)$ 中只含一个未知参数 θ,如果 θ 的充分估计量 $T(\xi_1,\cdots,\xi_n)$ 存在,且似然方程

$$\frac{\partial\ln L(\xi_1,\cdots,\xi_n;\theta)}{\partial\theta}=0 \qquad (7.3.3)$$

有解,则其解一定是 $t=T(x_1,\cdots,x_n)$ 的函数.

证 如果 T 是 θ 的充分估计量,则由(7.3.1)式知

$$L(\xi_1,\cdots,\xi_n;\theta)=\prod_{i=1}^{n}f(\xi_i;\theta)=f_T(t,\theta)h(\xi_1,\cdots,\xi_n),$$

且

$$\frac{\partial\ln L(\xi_1,\cdots,\xi_n;\theta)}{\partial\theta}=\frac{\partial\ln f_T(t,\theta)}{\partial\theta}.$$

因而 (7.3.3) 的解一定是 t 的函数, 即依赖于统计量 T 的观察值 t. ▌

这个定理表明,如果参数 θ 的充分估计存在,且似然方程有解并为 θ 的极大似然估计, 那么这个极大似然估计具有充分估计量的优良性.

定理 7.3.4 如果$T(\xi_1,\cdots,\xi_n)$是参数θ的优效估计量，那么$t=T(\xi_1,\cdots,\xi_n)$在满足$f_T(t,\theta)>0$的区域内，概率为1地都是似然方程的解.

证 由定理7.3.2知，$f_T(t,\theta)$满足(7.3.2)式，并有

$$\frac{\partial \ln L(\theta)}{\partial\theta}=C(\theta)[T-\theta], \tag{7.3.4}$$

所以统计量T是似然方程的解. ▌

这个定理表明，如果参数θ的优效估计存在，且似然方程有解并为θ的极大似然估计，那么这个极大似然估计具有优效估计的优良性.

定理 7.3.5(杜琪定理) 设总体ξ为连续型随机变数，密度函数为$f(x;\theta)$，假定

(1) 对任一数值x及$\theta\in A\subset\Omega$，导数$\frac{\partial \ln f}{\partial\theta}$，$\frac{\partial^2 \ln f}{\partial\theta^2}$及$\frac{\partial^3 \ln f}{\partial\theta^3}$都存在.

(2) 对任一$\theta\in A\subset\Omega$，有

$$0<I(\theta)=E\left[\frac{\partial \ln f(\xi,\theta)}{\partial\theta}\right]^2<\infty.$$

(3) 对每一个$\theta\in A$，关系式

$$\left|\frac{\partial f}{\partial\theta}\right|<F_1(x),\ \left|\frac{\partial^2 f}{\partial\theta^2}\right|<F_2(x)\text{及}\left|\frac{\partial^3 f}{\partial\theta^3}\right|<F_3(x)$$

成立，其中$F_1(x)$、$F_2(x)$及$F_3(x)$在$x\in R_1$上可积，而且$F_3(x)$满足条件$\int_{-\infty}^{\infty}F_3(x)f(x,\theta)dx<M$，$M$为与参数$\theta$无关的常数.

那么，参数θ的极大似然法估计量$\hat{\theta}(\xi_1,\cdots,\xi_n)$依概率收敛于$\theta$，而且$\sqrt{n}[\hat{\theta}(\xi_1,\cdots,\xi_n)-\theta]$的极限分布为正态$N\left(0,\frac{1}{\sqrt{I(\theta)}}\right)$.

定理证明略去，可参看[5]中第13章定理7—3.

这个定理指出，对于足够大的容量的子样，参数θ的极大似然

估计量 $\hat{\theta}(\xi_1,\cdots,\xi_n)$的概率分布，可用正态$N\left(\theta,\frac{1}{\sqrt{nI(\theta)}}\right)$ 作近似计算. 但是，还不能说 $\hat{\theta}$ 是 θ 的渐近优效估计，因为 $D(\hat{\theta})$ 及其极限都不一定存在.

在这一段中，首先看到常见的分布如泊松分布、均匀分布及正态分布、二项分布，其未知参数的极大似然法估计量都是充分估计量，接着证明了几个定理. 从几个定理的结论看到，优效估计量一定是充分估计量，极大似然法估计量一定是充分估计量的函数. 可以证明，充分估计量的单值函数也是一个充分估计量. 下面再介绍一个定理，这个定理阐明优效估计量与最优无偏估计量之间的关系.

定理 7.3.6 设总体 ξ 的分布函数为 $F(x;\theta)$，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为其子样，如果

(1) $\hat{\theta}$ 是 θ 的无偏估计量且方差有限，即对一切 $\theta\in\Omega$

$$E_\theta(\hat{\theta})=\theta,\qquad D_\theta(\hat{\theta})<\infty,$$

(2) T 是 θ 的充分统计量. 记

$$\hat{\theta}^*=E_\theta(\hat{\theta}|T),\tag{7.3.5}$$

它是 $\hat{\theta}$ 关于 T 的条件数学期望.

则有

(1) $\hat{\theta}^*$是不含有 θ 的并为 θ 的无偏估计量，即对一切 $\theta\in\Omega$，有 $E_\theta(\hat{\theta}^*)=\theta$;

(2) $D_\theta(\hat{\theta}^*)\leqslant D_\theta(\hat{\theta})$，

当且仅当 $P_\theta\{\hat{\theta}=\hat{\theta}^*\}=1$ 时，等号成立.

证 利用上述几个定理的结论，可以扼要地叙述这个定理的证明. 由于估计量 $\hat{\theta}$ 不带有未知参数 θ，子样在充分统计量 $T=t$ 的条件下的条件分布与 θ 无关，因而 $\hat{\theta}^*=E_\theta(\hat{\theta}|T)$ 不带有未知参数 θ，是一统计量，可作为 θ 估计量. 从第三章讲到的条件数学期

望性质知，对一切 $\theta\in\Omega$

$$E_\theta(\hat{\theta}^*)=E_\theta[E_\theta(\hat{\theta}|T)]=E_\theta(\hat{\theta})=\theta.$$

另一方面，

$$\begin{aligned}D_\theta(\hat{\theta})&=E_\theta[(\hat{\theta}-\hat{\theta}^*)+(\hat{\theta}^*-\theta)]^2\\&=E_\theta(\hat{\theta}-\hat{\theta}^*)^2+2E_\theta(\hat{\theta}-\hat{\theta}^*)(\hat{\theta}^*-\theta)+D_\theta(\hat{\theta}^*)\\&=E_\theta(\hat{\theta}-\hat{\theta}^*)^2+D_\theta(\hat{\theta}^*)\geqslant D_\theta(\hat{\theta}^*),\end{aligned}$$

其中等号成立的充要条件为

$$E_\theta(\hat{\theta}-\hat{\theta}^*)^2=0.$$

即有

$$P_\theta\{\hat{\theta}=\hat{\theta}^*\}=1.\ \blacksquare$$

(在推导上述不等式过程中，应用了关系式：

$$E_\theta(\hat{\theta}-\hat{\theta}^*)(\hat{\theta}^*-\theta)=0,$$

这个等式分别在连续型与离散型的情况下直接可得到证明).

二、完备性

由上所知，最优无偏估计量可局限在充分无偏估计类中寻找. 如果充分无偏估计量是唯一的，那么它就是最优无偏估计了. 试问，在什么情况下，它是唯一的呢？为此引进另一个重要的概念，即完备性.

定义 7.3.2(完备性统计量) 设总体 ξ 的分布函数为$F(x;\theta)$ $\theta\in\Omega$，$g(\xi)$为一随机变数，如果对一切 $\theta\in\Omega$，

$$E_\theta[g(\xi)]=0$$

成立时，对于一切 $\theta\in\Omega$ 必有：

$$P_\theta\{g(\xi)=0\}=1,$$

则称 $F(x,\theta)$是完备的. 若 $\xi_1,\cdots,\xi_n$ 为子样，统计量 $T(\xi_1,\cdots,\xi_n)$ 的分布函数是完备的，则称T 为完备统计量.

定理 7.3.7 设总体 ξ 的分布函数为 $F(x,\theta)$，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为其子样，$T(\xi_1,\cdots,\xi_n)$是 θ 的充分完备统计量. 如果 θ 的无偏估

计量存在，记为 $\hat{\theta}$，则 $\hat{\theta}^*=E(\hat{\theta}|T)$ 是唯一的最优无偏估计量.

证 设 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 是参数 θ 的任意两个无偏估计量，由定理 7.3.6 知，$E(\hat{\theta}_1|T)$ 及 $E(\hat{\theta}_2|T)$ 也是参数 θ 的无偏估计量，即对一切 $\theta\in\Omega$，有：

$$E_\theta[E(\hat{\theta}_1|T)]=E_\theta[E(\hat{\theta}_2|T)]=\theta.$$

而且对一切 $\theta\in\Omega$，有

$$D_\theta[E(\hat{\theta}_1|T)]\leqslant D_\theta(\hat{\theta}_1),$$

$$D_\theta[E(\hat{\theta}_2|T)]\leqslant D_\theta(\hat{\theta}_2).$$

由于对一切$\theta\in\Omega$，有：

$$E_\theta[E(\hat{\theta}_1|T)-E(\hat{\theta}_2|T)]=0,$$

且 T 是完备统计量，所以对一切 $\theta\in\Omega$，

$$P_\theta\{E(\hat{\theta}_1|T)=E(\hat{\theta}_2|T)\}=1.$$

因此参数 θ 的充分无偏估计量（在几乎处处相等的意义下）是唯一的. 由定理 7.3.6 知，若 $\hat{\theta}$ 为参数 θ 的无偏估计量，则 $\hat{\theta}^*=E(\hat{\theta}|T)$ 是参数 θ 的最优无偏估计量. ▌

对于一个估计量 $\hat{\theta}$，直接计算它的数学期望，可以判断它是否为参数 θ 的无偏估计量，但是要判断它是否为 θ 的充分完备估计量，一般是较麻烦的. 现在引进一类很重要的分布族，它包含了常用的一些分布，如泊松分布、正态分布、指数分布、二项分布、Γ 分布等，对这类分布族，寻找参数的充分完备估计量是方便的.

定义7.3.3（指数族分布） 设总体 ξ 的密度函数为 $f(x;\theta)$，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为其子样，子样的联合密度函数具有形式

$$f(X;\theta)=C(\theta)\exp\{b(\theta)T(X)\}h(X), \qquad (7.3.6)$$

其中 $X=(x_1,\cdots,x_n)$，$C(\theta)$、$b(\theta)$ 只与参数 θ 有关而与子样无关，$h(X)$、$T(X)$ 只与子样有关而不带未知参数 θ. 则称 $f(x;\theta)$ 为指数族分布. 对于离散型随机变数，如果它的概率函数 $P\{x;\theta\}$ 具有 (7.3.6) 式所定义的形式，也同样称它为指数族分布.

如果分布函数为 $F(x;\theta_1,\cdots,\theta_k)$，即有 k 维参数向量 $\theta=(\theta_1,\cdots,\theta_k)$，则(7.3.6)式中的指数部分为 $\exp\left\{\sum_{j=1}^{k}b_j(\theta)T_j(X)\right\}$ 形式，其中 $T_j(X)$ 为统计量，$j=1,\cdots,k$.

定理7.3.8 设总体 ξ 的密度函数 $f(x;\theta)$ 为指数族分布，$\xi_1,\cdots,\xi_n$ 为其子样，联合密度函数具有如下形式：

$$f(X;\theta)=C(\theta)\exp\left\{\sum_{j=1}^{k}b_j(\theta)T_j(X)\right\}h(X),$$

其中 $\theta=(\theta_1,\cdots,\theta_k)$，$X=(x_1,\cdots,x_n)$，$\theta\in\Omega$. 如果 Ω 中包含有一个 k 维矩形，而且 $B=(b_1,\cdots,b_k)$ 的值域包含有一个 k 维开集，则

$$\mathbf{T}(X)=(T_1(X),\cdots,T_k(X))$$

是 k 维参数向量 θ 的充分完备估计量.

证明请参看[2]的定理1.6.1.

例7.3.6 设总体 ξ 服从参数 $\lambda>0$ 的泊松分布 $P(x;\lambda)$，即有

$$P(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}\qquad(x=0,1,2,\cdots).$$

$\xi_1,\cdots,\xi_n$ 为其子样，它的联合概率函数为

$$P\{\xi_1=x_1,\cdots,\xi_n=x_n;\lambda\}=\frac{\lambda^{\sum_{i=1}^{n}x_i}e^{-n\lambda}}{\prod_{i=1}^{n}(x_i!)}$$

$$=e^{-n\lambda}\exp\left\{\sum_{i=1}^{n}x_i\cdot\ln\lambda\right\}\frac{1}{\prod_{i=1}^{n}(x_i!)}.$$

与(7.3.6)式对比，有

$$C(\lambda)=e^{-n\lambda};$$

$$h(X)=\frac{1}{\prod_{i=1}^{n}(x_i!)};$$

$$\mathrm{T}(X)=\bar{x}$$

$$b(\lambda)=n\ln\lambda.$$

因此，子样平均值$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$是参数$\lambda$的充分完备估计量．由于$\bar{\xi}$是$\lambda$的无偏估计量，因而$\bar{\xi}$是$\lambda$的最优无偏估计量．

例7.3.7 设总体ξ服从两点分布，即有：

$$P\{\xi=x\}=\begin{cases}p, & x=1,\\ 1-p, & x=0,\end{cases}$$

$0<p<1$．$\xi_1,\cdots,\xi_n$为其子样，它的联合概率函数为

$$\begin{aligned}&P\{\xi_1=x_1,\cdots,\xi_n=x_n;p\}\\&=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}\\&=(1-p)^n\left(\frac{p}{1-p}\right)^{\sum_{i=1}^{n}x_i}\\&=(1-p)^n\exp\left\{n\bar{x}\cdot\ln\left(\frac{p}{1-p}\right)\right\}.\end{aligned}$$

与(7.3.6)对比，有

$$C(p)=(1-p)^n;$$

$$h(X)=1;$$

$$b(p)=n\ln\left(\frac{p}{1-p}\right);$$

$$\mathrm{T}(X)=\bar{x}.$$

因此，子样平均值$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$是参数$p$的充分完备估计量．由于$\bar{\xi}$是$p$的无偏估计量，因而$\bar{\xi}$是$p$的最优无偏估计量．

例 7.3.8 设总体 ξ 服从均匀分布，其密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & 0<x<\theta,\\ 0, & \text{其它},\end{cases}$$

$\theta\in\Omega=(0,1)$，$\xi_1,\cdots,\xi_n$ 为其子样，它的联合密度函数为

$$f(x_1,\cdots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0<x_1,\cdots,x_n<\theta,\\ 0, & \text{其它}.\end{cases}$$

这时不便于与(7.3.6)式对比寻找参数 θ 的充分完备估计量. 我们在例 7.3.3 中已指出 $T(\xi_1,\cdots,\xi_n)=\max\limits_{1\leqslant i\leqslant n}\xi_i=\xi_n^*$ 是参数 θ 的充分估计量，下面证明它也是 θ 的完备估计量.

子样极大值 $\xi_n^*=\max\limits_{1\leqslant i\leqslant n}\xi_i$ 的分布函数为

$$F_n(t)=P\{\xi_n^*<t\}=[P\{\xi<t\}]^n$$

$$=\begin{cases}0 & t\leqslant 0\\ \left(\dfrac{t}{\theta}\right)^n & 0<t\leqslant\theta\\ 1 & t>\theta\end{cases}$$

密度函数为

$$f_T(t)=\begin{cases}\dfrac{nt^{n-1}}{\theta^n}, & 0<t<\theta,\\ 0, & \text{其它}.\end{cases}$$

如果有函数 $g(t)$ 对于一切 $0<\theta<1$，使得

$$\int_0^\theta g(t)f_T(t;\theta)dt=\frac{n}{\theta^n}\int_0^\theta g(t)t^{n-1}dt\equiv 0,$$

我们来证对一切 $0<\theta<1$，有：

$$g(t)\equiv 0.$$

记

$$h(\theta)=\int_0^\theta g(t)t^{n-1}dt=0,$$

则

$$h'(\theta)=g(\theta)\theta^{n-1}=0$$

几乎处处成立. 因为 $\theta\neq0$,从而推知对一切 $0<\theta<1$,

$$g(\theta)=0$$

于是 $g(t)$在$(0,1)$上几乎处处为零,因此 ξ_n^* 是参数 θ 的完备估计量.

但是,$\xi_n^*=\max\limits_{1\leq i\leq n}\xi_i$ 不是 θ的无偏估计量,而由$E(\xi_n^*)=\frac{n\theta}{n+1}$知,$\frac{n+1}{n}\xi_n^*$为 θ 的无偏估计量. 因而

$$\frac{n+1}{n}\xi_n^*=\frac{n+1}{n}\max_{1\leq i\leq n}\xi_i$$

为 θ 的最优无偏估计量.

例 7.3.9 设总体 ξ 服从正态 $N(a,\sigma)$,$\xi_1,\cdots,\xi_n$ 为其子样,它的联合密度函数为

$$\begin{aligned}f(x_1,\cdots,x_n)&=\frac{1}{(\sqrt{2\pi}\,\sigma)^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-a)^2\right\}\\&=\frac{1}{(2\pi\sigma^2)^{n/2}}e^{-\frac{na^2}{2\sigma^2}}\\&\quad\times\exp\left\{-\frac{n}{2\sigma^2}\left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)+\frac{na}{\sigma^2}\bar{x}\right\},\end{aligned}$$

与(7.3.6)式对比,$\theta=(a,\sigma^2)$,有

$$C(\theta)=\frac{1}{(2\pi\sigma^2)^{n/2}}e^{-\frac{na^2}{2\sigma^2}};$$

$$\mathbf{T}(X)=\left(\bar{x},\frac{1}{n}\sum_{i=1}^n x_i^2\right),$$

$$h(X)=1.$$

因此，子样平均值$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$及方差$S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$是$a$及$\sigma^2$的充分完备估计量，$\bar{\xi}$及$S^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2$为$a$及$\sigma^2$的最优无偏估计量.

§7.4 区间估计

设总体ξ的分布函数$F(x;\theta)$中的θ为未知参数，$\xi_1,\cdots,\xi_n$为其子样，现建立两个统计量$T_1(\xi_1,\cdots,\xi_n)$及$T_2(\xi_1,\cdots,\xi_n)$，并满足不等式：

$$T_1(\xi_1,\cdots,\xi_n)<T_2(\xi_1,\cdots,\xi_n),$$

则称$[T_1, T_2]$为随机区间. 随机区间的端点及区间长度都是子样的函数，因而都是统计量.

定义 7.4.1 （区间估计） 设α为一给定的常数，满足$0<\alpha<1$，如果关系式：

$$P\{T_1\leqslant\theta\leqslant T_2\}=1-\alpha \tag{7.4.1}$$

成立，并用这个随机区间作为参数θ的估计，则称$[T_1,T_2]$是参数θ的置信水平为$1-\alpha$的区间估计. 称α为显著性水平，称$T_2(\xi_1,\cdots,\xi_n)$、$T_1(\xi_1,\cdots,\xi_n)$分别为上、下置信限.

参数θ的区间估计的意义可解释为：建立统计量$T_1(\xi_1,\cdots,\xi_n)$及$T_2(\xi_1,\cdots,\xi_n)$后，T_1及T_2的观察值依赖于子样$\xi_1,\cdots,\xi_n$的观察值，现对子样作了N次观察，每次子样的观察值记为$(x_{1k},\cdots,x_{nk})$，$k=1,\cdots,N$，它对应于t_{1k}及t_{2k}观察值所组成的一个区间$[t_{1k},t_{2k}]$，这N个区间不一定都包含着参数θ的真值，有些区间包含着θ的真值而另一些区间不包含它，当(7.4.1)式成立时，随机区间$[T_1,T_2]$包含着参数θ真值的频率近似为$1-\alpha$. 此时，若认为

"区间$[T_1,T_2]$包含着参数θ的真值",那么这种"认为"犯错误的概率为α. 也就是说,随机区间$[T_1,T_2]$以$1-\alpha$的概率包含着参数θ,这是参数θ的区间估计的本质所在. 注意,我们并非说参数θ以$1-\alpha$的概率落入随机区间$[T_1,T_2]$,因为参数θ是非随机的.

例 7.4.1 设总体ξ服从$N(a,\sigma)$,$\xi_1,\cdots,\xi_n$为其子样,试求a及σ^2的区间估计.

解 先考虑a的区间估计:

(1) 在方差$\sigma^2=\sigma_0^2$为已知时,求a的区间估计.

我们知道,$\bar{\xi}$服从正态$N\left(a,\frac{\sigma_0}{\sqrt{n}}\right)$,因而统计量

$$U=\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}}$$

服从正态$N(0,1)$. 给定置信水平$1-\alpha$,在标准正态分布表中查得临界值u_α,使得

$$P\left\{\left|\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}}\right|\leqslant u_\alpha\right\}=1-\alpha,$$

亦即

$$P\left\{\bar{\xi}-u_\alpha\frac{\sigma_0}{\sqrt{n}}\leqslant a\leqslant\bar{\xi}+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right\}=1-\alpha. \qquad (7.4.2)$$

就是说,随机区间$\left[\bar{\xi}-u_\alpha\frac{\sigma_0}{\sqrt{n}},\bar{\xi}+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right]$是数学期望$a$的置信水平为$1-\alpha$的区间估计.

对于不同的置信水平$1-\alpha$,a的区间估计随之不同. 如表7.4.1所示

(2) 在方差σ^2未知时,求a的区间估计.

当方差σ^2未知时,可用子样方差S^2估计σ^2. 由(6.3.7)式知道,

$$\sqrt{n-1}\frac{\bar{\xi}-a}{S}\sim t_{(n-1)}.$$

表 7.4.1

α	u_α	$\left[\bar\xi-u_\alpha\frac{\sigma_0}{\sqrt{n}},\bar\xi+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right]$
0.01	2.58	$\left[\bar\xi-2.58\frac{\sigma_0}{\sqrt{n}},\bar\xi+2.58\frac{\sigma_0}{\sqrt{n}}\right]$
0.05	1.96	$\left[\bar\xi-1.96\frac{\sigma_0}{\sqrt{n}},\bar\xi+1.96\frac{\sigma_0}{\sqrt{n}}\right]$
0.10	1.65	$\left[\bar\xi-1.65\frac{\sigma_0}{\sqrt{n}},\bar\xi+1.65\frac{\sigma_0}{\sqrt{n}}\right]$

给定置信水平 $1-\alpha$, $0<\alpha<1$, 在 t 分布表中查自由度 $n-1$ 对应的临界值 $t_{(n-1)}(\alpha)$, 使得

$$P\left\{\left|\sqrt{n-1}\frac{\bar\xi-a}{S}\right|\geqslant t_{(n-1)}(\alpha)\right\}=\alpha,$$

亦即

$$P\left\{\bar\xi-t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\leqslant a\leqslant\bar\xi+t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\right\}=1-\alpha. \tag{7.4.3}$$

就是说, $\left[\bar\xi-t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}},\ \bar\xi+t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\right]$ 是 a 的置信水平为 $1-\alpha$ 的区间估计.

现考虑方差 σ^2 的区间估计:

由第六章的定理 6.3.1 知, 若总体 ξ 服从正态 $N(a, \sigma)$, 则有:

$$\frac{nS^2}{\sigma^2}\sim\chi^2{}_{(n-1)}.$$

给定置信水平 $1-\alpha$, $0<\alpha<1$, 在 χ^2-分布临界值表中查自由度 $n-1$ 对应的两个数值 $\chi^2{}_{(n-1)}\left(\frac{\alpha}{2}\right)$ 及 $\chi^2{}_{(n-1)}\left(1-\frac{\alpha}{2}\right)$使得

$$P\left\{\frac{nS^2}{\sigma^2}\geqslant\chi^2{}_{(n-1)}\left(\frac{\alpha}{2}\right)\right\}=\frac{\alpha}{2};$$

$$P\left\{\frac{nS^2}{\sigma^2} \geqslant \chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)\right\} = 1-\frac{\alpha}{2}.$$

即有

$$P\left\{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right) \leqslant \frac{nS^2}{\sigma^2} \leqslant \chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)\right\} = 1-\alpha;$$

$$P\left\{\frac{nS^2}{\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)} \leqslant \sigma^2 \leqslant \frac{nS^2}{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)}\right\} = 1-\alpha. \qquad (7.4.4)$$

就是说，$\left[\frac{nS^2}{\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)}, \frac{nS^2}{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)}\right]$为方差$\sigma^2$的置信水平$1-\alpha$的区间估计.

上述的区间估计是在总体ξ服从正态$N(a, \sigma)$的情况下讨论的. 当总体ξ不服从正态时，若其数学期望$E(\xi)$及方差$D(\xi)$都有限，则由第五章的定理 5.4.3 知，统计量$\eta_n=\frac{\bar{\xi}-a}{\sigma/\sqrt{n}}$的极限分布为正态$N(0,1)$，即当$n$足够大时，子样平均$\bar{\xi}$渐近服从正态$N\left(a, \frac{\sigma}{\sqrt{n}}\right)$. 此时，可应用(7.4.2)近似地给出$a$的区间估计. 如果方差$\sigma^2$未知，要用子样方差$S^2$估计$\sigma^2$，则可应用(7.4.3)式近似地给出$a$的区间估计.

下面的例子，则不必用(7.4.2)及(7.4.3)两式作近似计算，它可直接导出子样平均$\bar{\xi}$的精确分布，从而求出参数的区间估计.

例 7.4.2 设总体ξ服从指数分布，密度函数为：

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\ \lambda>0; \\ 0, & 其它,\end{cases}$$

求参数λ的置信水平$1-\alpha$的上置信限.

解 我们知道，ξ的特征函数为

$$\varphi(t)=\left(1-j\frac{t}{\lambda}\right)^{-1},$$

于是，子样平均 $\bar{\xi}$ 的特征函数为

$$\varphi_{\bar{\xi}}(t)=\left(1-j\frac{t}{n\lambda}\right)^{-n}.$$

记 $\eta=2n\lambda\bar{\xi}$，则 η 的特征函数为

$$\varphi_{\eta}(t)=(1-2jt)^{-n}.$$

可见，统计量 η 服从自由度为 $2n$ 的 χ^2 分布. 给定置信水平 $1-\alpha$，查得临界值 $\chi_{2n}^{2}(\alpha)$，使得

$$P\{2n\lambda\bar{\xi}\leqslant\chi_{2n}^{2}(\alpha)\}=1-\alpha,$$

亦即

$$P\left\{0<\lambda\leqslant\frac{1}{2n\bar{\xi}}\chi_{2n}^{2}(\alpha)\right\}=1-\alpha, \tag{7.4.5}$$

因此 $\left(0,\ \frac{1}{2n\bar{\xi}}\chi_{2n}^{2}(\alpha)\right)$ 是参数 λ 的置信水平为 $1-\alpha$ 的区间估计.

也可以这样选取 λ 的区间估计:

$$P\left\{2n\lambda\bar{\xi}\geqslant\chi_{2n}^{2}\left(\frac{\alpha}{2}\right)\right\}=\frac{\alpha}{2};$$

$$P\left\{2n\lambda\bar{\xi}\geqslant\chi_{2n}^{2}\left(1-\frac{\alpha}{2}\right)\right\}=1-\frac{\alpha}{2};$$

$$P\left\{\chi_{2n}^{2}\left(1-\frac{\alpha}{2}\right)\leqslant 2n\lambda\bar{\xi}\leqslant\chi_{2n}^{2}\left(\frac{\alpha}{2}\right)\right\}=1-\alpha;$$

$$P\left\{\frac{1}{2n\bar{\xi}}\chi_{2n}^{2}\left(1-\frac{\alpha}{2}\right)\leqslant\lambda\leqslant\frac{1}{2n\bar{\xi}}\chi_{2n}^{2}\left(\frac{\alpha}{2}\right)\right\}=1-\alpha.$$

*§7.5　极大极小估计与容许估计

一、决策论的基本概念

在这节里，我们将运用更一般的观点——决策论来考察点估计问题. 为此，我们先介绍决策论的一些名词和概念，如决策函数、风险函数等.

设总体 ξ 的分布函数为 $F(x;\theta)\theta\in\Omega$，$\theta$ 为未知参数，Ω 为参

数空间，$\xi_1, \cdots, \xi_n$ 为其子样．求参数 θ 的点估计，可看作为由子样 $\xi_1, \cdots, \xi_n$ 的一次观察值——子样空间 $\mathscr{X}$ 中的一点 $(x_1, \cdots, x_n)$，对未知参数 θ 采取一种决定．在数理统计学中，我们称所采取的一个决定为决策，将可能采取的全部决策所组成的集合称为决策空间，记为 $\mathscr{A}$．因此，对未知参数 θ 或对分布 $F(x;\theta)$ 作出估计，实质上是对子样空间 $\mathscr{X}$ 中的每一点 $(x_1, \cdots, x_n)$，在决策空间 $\mathscr{A}$ 中寻找一点 $a \in \mathscr{A}$ 与之对应，于是我们要选取一个在决策空间 $\mathscr{A}$ 中取值的子样 $\xi_1, \cdots, \xi_n$ 的函数 $d(\xi_1, \cdots, \xi_n)$．当子样观察值为 $(x_1, \cdots, x_n)$ 时，我们采取决策 $a = d(x_1, \cdots, x_n)$，$a \in \mathscr{A}$，$d(\xi_1, \cdots, \xi_n)$ 是一个统计量，我们称它为决策函数或判决函数．

例 7.5.1 点估计问题的决策论解释

设总体 ξ 有分布函数 $F(x;\theta)$，$\theta \in \Omega$，θ 为未知参数．要求由子样 $\xi_1, \cdots, \xi_n$ 建立统计量作为 θ 的估计量．这时，决策空间 $\mathscr{A} = \Omega$，即与参数空间 Ω 重合，而每个估计量就是一个决策函数．

我们知道，对于每个具体的统计问题，总存在着许多不同的决策函数可利用，因而就需要发展某些理论，能够使我们评价决策函数的好坏，从而选择一个好的决策函数．为此引入损失函数，记为 $L(\theta, a)$，它是参数 θ 及决策 a 的一个实值非负函数，它表示参数为 θ 时，而采取决策为 a 所造成的损失．当然，我们希望当其参数为 θ 时，能采取一个"最好"的决策，即使损失 L 达到最小的决策．但是 θ 的参数值是未知的，此时我们就无法知道最好的决策是什么．我们是利用子样建立决策函数 $d(\xi_1, \cdots, \xi_n)$ 来产生决策的，它所对应的损失为 $L(\theta, a) = L[\theta, d(\xi_1, \cdots, \xi_n)]$，所以这个损失是依赖于子样的，因而它是一个统计量．因此，我们还不能运用基于子样的某次观察 $(x_1, \cdots, x_n)$ 所采取的某个决策所带来的损失 $L[\theta, d(x_1, \cdots, x_n)]$ 来衡量，而应从总体上来评价，现采用平均损失来作评价．记

$$R(\theta, d)=E_\theta\{L[\theta, d(\xi_1, \cdots, \xi_n)]\}. \qquad (7.5.1)$$

它是损失函数在总体 ξ 的分布取参数为 θ 时计算得到的数学期望．称 $R(\theta, d)$ 为风险函数．它是参数 θ 及决策函数 $d(\xi_1, \cdots, \xi_n)$ 的函数，它表示参数为 θ 时而采取决策 $d(\xi_1, \cdots, \xi_n)$ 所造成的平均损失.

例 7.5.2 设总体 ξ 服从泊松分布 $P(x, \lambda)$，$\lambda\in\Omega=(0, \infty)$，要求估计未知参数 λ. 如果我们选取损失函数为

$$L(\lambda, d)=(d-\lambda)^2, \qquad (7.5.2)$$

即二次损失函数．则对 λ 的任一估计量 $d(\xi_1, \cdots, \xi_n)$，风险函数是

$$R(\lambda, d)=E_\lambda[L(\lambda, d)]=E_\lambda[(d-\lambda)^2].$$

如进一步要求估计量 $d(\xi_1, \cdots, \xi_n)$ 是无偏的，即

$$E_\lambda[d(\xi_1, \cdots, \xi_n)]=\lambda, \qquad (7.5.3)$$

则风险函数是 $R(\lambda, d)=E_\lambda[d-E_\lambda(d)]^2=D_\lambda(d). \qquad (7.5.4)$

它是决策函数 $d(\xi_1, \cdots, \xi_n)$ 的方差．例如取 $d(\xi_1, \cdots, \xi_n)=\bar{\xi}$，那么所对应的风险函数为：

$$R(\lambda, d)=D_\lambda(\bar{\xi})=\frac{\lambda}{n}.$$

若如取 $d(\xi_1, \cdots, \xi_n)=\xi_1$，那么所对应的风险函数为：

$$R(\lambda, d)=D_\lambda(\xi_1)=\lambda.$$

显然，当 $n>1$ 时，后者的"风险"比前者大，称 $\bar{\xi}$ 优于 ξ_1.

二、极大极小估计

由上面可看到风险函数提供了一个衡量决策函数好坏的准则．当然我们希望选取一个决策函数，使得它的风险尽可能地小.

定义 7.5.1 极大极小估计．设 $\mathscr{D}$ 是由决策函数为元素组成的集合，若有 $d^*(\xi_1, \cdots, \xi_n)\in\mathscr{D}$，对任一 $d(\xi_1, \cdots, \xi_n)\in\mathscr{D}$，使得

$$\sup_{\theta\in\Theta}\{R(\theta, d^*)\}\leqslant\sup_{\theta\in\Theta}\{R(\theta, d)\}, \qquad (7.5.5)$$

则称 $d^*(\xi_1, \cdots, \xi_n)$ 为参数 θ 的极大极小估计量.

这个定义的直观背景是很有意思的. 使得最大风险达到最小的决策函数$d^*(\xi_1,\cdots,\xi_n)$是考虑到最不利的情况并要求最不利的情况尽可能地好的一种策略，这就是通常所说的从最坏处着想而争取最好的结果，是一种出于稳妥的考虑，也是一种偏于保守的考虑.

例 7.5.3 设总体ξ服从两点分布，即有:

$$P\{\xi=x\}=\begin{cases}p\quad, & x=1;\\1-p, & x=0,\end{cases}$$

其中$p\in\Omega=\left\{\frac{1}{4},\frac{1}{2}\right\}$，试求参数$p$的极大极小估计量.

解 在本例中参数的决策空间$\mathscr{A}=\left\{\frac{1}{4},\frac{1}{2}\right\}$. 设损失函数$L(p,a)$由表(7.5.1)给定.

表 7.5.1

$L(p,a)$ (a值 / p值)	$a_1=\frac{1}{4}$	$a_2=\frac{1}{2}$
$p_1=\frac{1}{4}$	1	4
$p_2=\frac{1}{2}$	3	2

如果我们选取容量为 1 的子样记为ξ_1，由于ξ_1仅取两个可能值及$\mathscr{A}$中只有两个元素，因而决策函数的集合$\mathscr{D}$是由 4 个元素所组成，分别以d_1,d_2,d_3,d_4表示，即有:

表 7.5.2

ξ \ $\mathscr{D}$	d_1	d_2	d_3	d_4
0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$

它们分别造成的风险 $R(p,a)$ 记录在表 7.5.3 中

表 7.5.3

d_i	$R(p_1,d_i)$	$R(p_2,d_i)$	$\sup R(p_i,d_i)$
d_1	1	3	3
d_2	4	2	4
d_3	$\frac{7}{4}$	$\frac{5}{2}$	$\frac{5}{2}$
d_4	$\frac{13}{4}$	$\frac{5}{2}$	$\frac{13}{4}$

由(7.5.1)式计算得到 $R(p,d)$，例如

$$\begin{aligned} R(p_1,d_3) &= E_{p_1}[L(p_1,d_3)] \\ &= L(p_1,a_1)\cdot P_{p_1}\{\xi=0\}+L(p_1,a_2)\cdot P_{p_1}\{\xi=1\} \\ &= 1\cdot(1-p_1)+4\cdot p_1=\frac{7}{4}, \end{aligned}$$

$$\begin{aligned} R(p_2,d_3) &= L(p_2,a_1)\cdot P_{p_2}\{\xi=0\}+L(p_2,a_2)\cdot P_{p_2}\{\xi=1\} \\ &= 3\cdot(1-p_2)+2\cdot p_2=\frac{5}{2}. \end{aligned}$$

由(7.5.5)式知，参数 p 的极大极小估计为：

$$\hat{p}(\xi_1)=d_3=\begin{cases}\frac{1}{4}, & \xi_1=0, \\ \frac{1}{2}, & \xi_1=1.\end{cases} \tag{7.5.6}$$

三、容许估计

定义 7.5.2 称 $d_1\in\mathscr{D}$ 处处不比 $d_2\in\mathscr{D}$ 差，如果对于每一 $\theta\in\Omega$，有

$$R(\theta,d_1)\leqslant R(\theta,d_2). \tag{7.5.7}$$

如果进一步假设对于某些 $\theta\in\Omega$，(7.5.7)中严格不等号成立，则称 d_1 优于 d_2。

称$d_0(\xi_1,\cdots,\xi_n)\in\mathscr{D}$是容许估计，如果不存在估计量$d(\xi_1,\cdots,\xi_n)$优于$d_0(\xi_1,\cdots,\xi_n)$，也就是说，所有处处不比$d$差的估计量都应与$d_0$等效，即风险函数相等。

例 7.5.4 设总体ξ服从正态$N(a,\sigma)$，$\xi_1,\cdots,\xi_n$为子样，$n\geqslant 2$. 要估计方差σ^2，选取平方损失函数

$$L(\sigma^2,\ d)=\left(\frac{d}{\sigma^2}-1\right)^2. \tag{7.5.8}$$

现在考察三个估计量:

$$d_1=\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2,\quad R(\sigma^2,\ d_1)=\frac{2}{n-1};$$

$$d_2=\hat{\sigma}_2^2=\frac{1}{n+1}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2, R(\sigma^2,\ d_2)=\frac{2}{n+1};$$

$$d_3=\hat{\sigma}_3^2=\frac{1}{n+2}\sum_{i=1}^{n}\xi_i^2, R(\sigma^2 d_3)=\frac{2}{n+2}+\frac{\mu^4}{\sigma^4}\left(\frac{n}{n+2}\right)^2.$$

记$\lambda=\mu^4/\sigma^4$，则有

$$R(\sigma^2,d_1)>R(\sigma^2,d_2)\text{，对一切 }\sigma^2>0;$$

$$R(\sigma^2,d_3)\leqslant R(\sigma^2,d_2)\text{，对 }\lambda\leqslant\frac{2(n+2)}{n^2(n+1)};$$

$$R(\sigma^2,d_3)>R(\sigma^2,\ d_2)\text{，对 }\lambda>\frac{2(n+2)}{n^2(n+1)}.$$

由例7.2.4知，$d_1(\xi_1,\cdots,\xi_n)$是σ^2的最优无偏估计. 但在(7.5.8)平方损失函数情况下，$d_2(\xi_1,\cdots,d_n)$优于$d_1(\xi_1,\cdots,\xi_n)$. 在参考书目[19]中还证明了$d_2(\xi_1,\cdots,\xi_n)$是σ^2的极大极小估计. 当$\lambda=\mu^4/\sigma^4$不超过某个数值时，$d_3(\xi_1,\cdots,\xi_n)$比$d_2(\xi_1,\cdots,\xi_n)$还优. 但是当λ超过某个数值时，d_3比d_2还差，甚至当$\lambda\to\infty$时，$R(\sigma^2,\ d_3)\to\infty$，因而看来$d_3(\xi_1,\cdots,\xi_n)$不是一个好的估计. 然而可以证明，$d_3$是$\sigma^2$的容许估计，而$d_2$却不是$\sigma^2$的容许估计.

上面分析说明，单纯从容许性角度来选择估计是不尽合理的，最常用的是极大极小性与容许性同时成立的估计量.

定理 7.5.1 (1) 若 d_0 是关于损失函数 $L(\theta, d)$ 基本唯一的极大极小估计，则 d_0 是 θ 的容许估计.

(2) 若 d_0 是容许的，而且在 $\theta\in\Omega$ 上为常数风险，则 d_0 是极大极小估计.

证 (1) 反证法：若 d_0 不是容许估计，则存在另一 d_1，d_1 优于 d_0，且 $d_1 \neq d_0$. 然而 d_1 也是 θ 的极大极小估计，这就与 d_0 的基本唯一性相矛盾.

(2) 若 d_0 不是极大极小估计，则存在 d_1，使得对每一 $\theta\in\Omega$，有

$$R(\theta, d_1) \leqslant \sup_{\theta\in\Omega}\{R(\theta, d_1)\}$$
$$< \sup_{\theta\in\Omega}\{R(\theta, d_0)\} = \text{常数}.$$

这与 d_0 为容许估计的假定相矛盾.

较详细的内容，可参看[19]的第八章.

*§7.6 贝叶斯估计

前述各节都是在未知参数 θ 作为非随机变数的情况下讨论参数估计问题，事前对未知参数 θ 没有什么信息. 但在有些情况下，事前可以提供未知参数 θ 的某些附加信息，这对于参数 θ 的估计是有益的. 如例 7.5.3 中，若知道 p_1 值比 p_2 有较大的可能成为参数 p 的真值，此时可以将参数 p 看作为一随机变数，具有某种概率分布，例如

$$P\{p=p_2\}=\frac{2}{3}, \qquad P\{p=p_1\}=\frac{1}{3}.$$

试问怎样利用关于参数 p 的附加信息来建立参数 p 的估计量?

下面引进贝叶斯方法. 较详细的内容，参看[19]的第四章.

设总体 ξ 的分布函数为 $F(x;\theta)$，$\theta\in\Omega$. 假定参数 θ 在参数空间 Ω 上具有给定的分布函数 $\pi(\theta)$，通常称 $\pi(\theta)$ 为 Ω 上 θ 的先验分布. 这时，用 $F(x|\theta)$ 表示给定 $\theta\in\Omega$ 时总体 ξ 的条件分布，而 $F(x,\theta)$ 就表示 ξ 及 θ 的联合分布函数，即有：

$$F(x,\theta)=\pi(\theta)F(x|\theta). \tag{7.6.1}$$

对于决策函数 $d(\xi_1,\cdots,\xi_n)$，由(7.5.1)式给出的风险函数 $R(\theta,d)$ 应写为：

$$R(\theta,d)=E\{L[\theta,d(\xi_1,\cdots,\xi_n)]|\theta\}, \tag{7.6.2}$$

它表示给定 $\theta\in\Omega$ 时的 $d(\xi_1,\cdots,\xi_n)$ 的条件平均损失，它依赖于 θ，因而 $R(\theta,d)$ 也是一随机变数，于是再把它关于 θ 进行平均，得到

$$\begin{aligned}B(d)&=E\{R[\theta,d(\xi_1,\cdots,\xi_n)]\}\\&=\int_\Omega E\{L[\theta,d(\xi_1,\cdots,\xi_n)]|\theta\}\pi(\theta)d\theta\end{aligned} \tag{7.6.3}$$

称 $B(d)$ 为决策函数 $d(\xi_1,\cdots,\xi_n)$ 的贝叶斯风险.

定义 7.6.1 (贝叶斯估计) 设总体 ξ 的分布函数 $F(x,\theta)$ 中参数 θ 为随机变数，对于任一决策函数 $d(\xi_1,\cdots,\xi_n)$，若有一决策函数 $d^*(\xi_1,\cdots,\xi_n)$，使得

$$B(d^*)=\min_{d\in\mathscr{D}}\{B(d)\}, \tag{7.6.4}$$

则称 d^* 为参数 θ 的贝叶斯估计量.

从(7.6.4)看到，贝叶斯估计量 $d^*(\xi_1,\cdots,\xi_n)$，是使得贝叶斯风险 $B(d)$ 达到最小的决策函数. 贝叶斯估计量是依赖于先验分布 $\pi(\theta)$ 的，也就是说，对应于不同的 $\pi(\theta)$，θ 的贝叶斯估计量是不同的.

下面对总体 ξ 为连续型随机变数的情形讨论两个定理，定理的结论也适用于总体 ξ 为离散型的情形，只是要用条件概率函数 $P(x|\theta)$ 代替条件密度函数 $f(x|\theta)$ 的位置，再用求和式代替积分式子.

由贝叶斯公式知道，有：

$$f(x,\theta)=\pi(\theta)f(x|\theta)=g(x)h(\theta|x);$$

$$h(\theta|x)=\frac{\pi(\theta)f(x|\theta)}{g(x)}, \qquad (7.6.5)$$

其中 $g(x)$ 是 ξ 的边沿分布密度函数，$h(\theta|x)$ 为给定 $\xi=x$ 时 θ 的条件分布函数，又称为 θ 的后验分布（相对于 $\pi(\theta)$ 而言，$\pi(\theta)$ 是在抽样试验以前就已知的，而 $h(\theta|x)$ 是在通过抽样试验以后确定的）。于是(7.6.3)式可写为：

$$B(d)=\underbrace{\int\cdots\int}_{(n+1)\text{重}}L[\theta,d(x_1,\cdots,x_n)]f(x_1,\cdots,x_n|\theta)$$

$$\times\pi(\theta)dx_1\cdots dx_nd\theta$$

$$=\underbrace{\int\cdots\int}_{n\text{重}}g(x_1,\cdots,x_n)\left\{\int_\Omega L[\theta,d(x_1,\cdots,x_n)]\right.$$

$$\left.\times h(\theta|x_1,\cdots,x_n)d\theta\right\}dx_1\cdots dx_n. \qquad (7.6.6)$$

定理 7.6.1 如果损失函数取二次式：

$$L(\theta,d)=[\theta-d(\xi_1,\cdots,\xi_n)]^2, \qquad (7.6.7)$$

则参数 θ 的贝叶斯估计量为：

$$d(\xi_1,\cdots,\xi_n)=E(\theta|\xi_1,\cdots,\xi_n)$$

$$=\int_\Omega\theta h(\theta|\xi_1,\cdots,\xi_n)d\theta. \qquad (7.6.8)$$

证 我们知道，

$$B(d)=\underbrace{\int\cdots\int}_{n\text{重}}g(x_1,\cdots,x_n)\left\{\int_\Omega[\theta-d(x_1,\cdots,x_n)]^2\right.$$

$$\left.\times h(\theta|x_1,\cdots,x_n)d\theta\right\}dx_1\cdots dx_n. \qquad (7.6.9)$$

(7.6.9)式对于 $d\in\mathscr{D}$ 中取最小值，几乎处处等价于下面的(7.6.10)式对于 $d\in\mathscr{D}$ 中取最小值：

$$\int_{\Omega}[\theta-d(x_1,\cdots,x_n)]^2h(\theta|x_1,\cdots,x_n)d\theta. \tag{7.6.10}$$

为方便书写，用X表示$(x_1,\cdots,x_n)$，(7.6.10)式可写为：

$$\begin{aligned}
&\int_{\Omega}[\theta-d(X)]^2h(\theta|X)d\theta \\
&=\int_{\Omega}[\theta-E(\theta|X)+E(\theta|X)-d(X)]^2h(\theta|X)d\theta \\
&=\int_{\Omega}[\theta-E(\theta|X)]^2h(\theta|X)d\theta \\
&\quad+\int_{\Omega}[E(\theta|X)-d(X)]^2h(\theta|X)d\theta \\
&\quad+2\int_{\Omega}[\theta-E(\theta|X)][E(\theta|X)-d(X)]h(\theta|X)d\theta,
\end{aligned} \tag{7.6.11}$$

其中

$$E(\theta|X)=\int_{\Omega}\theta h(\theta|X)d\theta.$$

又

$$\begin{aligned}
&\int_{\Omega}[\theta-E(\theta|X)][E(\theta|X)-d(X)]h(\theta|X)d\theta \\
&=\int_{\Omega}[\theta-E(\theta|X)]h(\theta|X)d\theta\cdot[E(\theta|X)-d(X)] \\
&=[E(\theta|X)-E(\theta|X)]\cdot[E(\theta|X)-d(X)]=0,
\end{aligned}$$

所以

$$\begin{aligned}
\int_{\Omega}[\theta-d(X)]^2h(\theta|X)d\theta&=\int_{\Omega}[\theta-E(\theta|X)]^2h(\theta|X)d\theta \\
&\quad+\int_{\Omega}[E(\theta|X)-d(X)]^2h(\theta|X)d\theta.
\end{aligned} \tag{7.6.12}$$

上式右边第一项与$d(X)$无关，显然当

$$d(X)=E(\theta|X)$$

几乎处处成立时，左边达到最小. ∎

例 7.6.1 设总体 ξ 服从两点分布，即有：

$$P\{x, p\}=\begin{cases}p & , x=1;\\ 1-p, & x=0,\end{cases}$$

$p\in\Omega=(0,1)$, ξ_1, $\cdots$, ξ_n 为其子样. 假定损失函数是二次误差，$L(p,d)=(p-d)^2$，并假定参数 p 是随机变数，服从均匀分布，即有：

$$\pi(p)=1, \qquad 0<p<1,$$

试求参数 p 的贝叶斯估计量.

解 子样 $(\xi_1,\cdots,\xi_n)$ 的边沿分布为:

$$g(x_1,\cdots,x_n)=\int_0^1 p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}dp$$

$$=B\left(\sum_{i=1}^n x_i+1, n+1-\sum_{i=1}^n x_i\right)$$

$$=\frac{\left[\left(\sum_{i=1}^n x_i\right)!\right]\left[\left(n-\sum_{i=1}^n x_i\right)!\right]}{(n+1)!},$$

参数 p 的后验分布为:

$$h(p|x_1,\cdots,x_n)=[(n+1)!]\frac{p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}}{\left[\left(\sum_{i=1}^n x_i\right)!\right]\left[\left(n-\sum_{i=1}^n x_i\right)!\right]}.$$

由定理 7.6.1 知，参数 p 的条件数学期望为:

$$p(x_1,\cdots,x_n)=\int_0^1 ph(p|x_1,\cdots,x_n)dp$$

$$=\frac{(n+1)!}{\left[\left(\sum_{i=1}^n x_i\right)!\right]\left[\left(n-\sum_{i=1}^n x_i\right)!\right]}$$

$$\times \int_0^1 p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} dp$$

$$= \frac{1+\sum_{i=1}^n x_i}{n+2},$$

也即参数 p 的贝叶斯估计量为

$$\hat{p}(\xi_1, \cdots, \xi_n) = \frac{1}{n+2}\left[1+\sum_{i=1}^n \xi_i\right]. \qquad (7.6.13)$$

例 7.6.2 设总体 ξ 服从正态 $N(a,1)$, $a \in \Omega = R_1$, $\xi_1, \cdots, \xi_n$ 为其子样. 假定参数 a 服从正态 $N(0,1)$, 又假定损失函数为二次误差 $L(a,d)=(a-d)^2$. 试求参数 a 的贝叶斯估计量.

解 子样的条件分布密度为

$$f(x_1, \cdots, x_n | a) = \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i - a)^2\right\},$$

子样和参数 a 的联合分布密度为

$$f(x_1, \cdots, x_n; a) = f(x_1, \cdots, x_n | a)\pi(a)$$

$$= \frac{1}{(\sqrt{2\pi})^{n+1}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (x_i - a)^2 - \frac{1}{2}a^2\right\},$$

子样的边沿分布密度为

$$g(x_1, \cdots, x_n) = \int_{-\infty}^{\infty} f(x_1, \cdots, x_n | a)\pi(a) da$$

$$= \frac{1}{(\sqrt{2\pi})^{n+1}} \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}$$

$$\times \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}[(n+1)a^2 - 2an\bar{x}]\right\} da$$

$$= \frac{1}{(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^n x_i^2 - \frac{n^2\bar{x}^2}{n+1}\right]\right\}$$

$$\times\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{n+1}{2}\left[a-\frac{n\bar{x}}{n+1}\right]^2\right\}da$$

$$=\frac{1}{\sqrt{n+1}(\sqrt{2\pi})^n}\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}x_i^2-\frac{n^2\bar{x}^2}{n+1}\right]\right\}.$$

于是参数 a 的后验分布为

$$h(a|x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n|a)\pi(a)}{g(x_1,\cdots,x_n)}$$

$$=\frac{\sqrt{n+1}}{\sqrt{2\pi}}\exp\left\{-\frac{n+1}{2}\left[a-\frac{n\bar{x}}{n+1}\right]^2\right\}.$$

由定理 7.6.1 知，参数 a 的条件数学期望为：

$$E(a|x_1,\cdots,x_n)=\int_{-\infty}^{\infty}ah(a|x_1,\cdots,x_n)da$$

$$=\int_{-\infty}^{\infty}\frac{a}{\sqrt{2\pi}\sigma_0}\exp\left\{-\frac{1}{2\sigma_0^2}\left[a-\frac{n\bar{x}}{n+1}\right]^2\right\}da$$

$$=\frac{n\bar{x}}{n+1},$$

其中$\sigma_0=\frac{1}{\sqrt{n+1}}$. 参数 a 的贝叶斯估计量为

$$\hat{a}(\xi_1,\cdots,\xi_n)=\frac{1}{n+1}\sum_{i=1}^{n}\xi_i.\qquad(7.6.14)$$

定理 7.6.2 设总体 ξ 的分布密度函数为 $f(x;\theta)$，$\theta\in\Omega$，$\xi_1,\cdots,\xi_n$ 为其子样，$d^*(\xi_1,\cdots,\xi_n)$是对应于先验分布 $\pi(\theta)$的参数 θ 的贝叶斯估计量. 如果风险函数 $R(\theta,d^*)$ 在 Ω 上是常数，则 d^* 也是 θ 的极大极小估计量.

证 设 d 是任一决策函数，由于 d^*是贝叶斯估计量，因此

$$B(d^*)\leqslant B(d).$$

由(7.6.3)式知

$$B(d^*)=\int_{\Omega}R(\theta,d^*)\pi(\theta)d\theta$$

$$\leqslant \int_{\Omega} R(\theta, d)\pi(\theta)d\theta = B(d),$$

因 $R(\theta, d^*)$ 在 Ω 上是常数，则有

$$\sup_{\theta\in\Omega}\{R(\theta, d^*)\} = R(\theta, d^*) = \int_{\Omega} R(\theta, d^*)\pi(\theta)d\theta$$

$$\leqslant \int_{\Omega} R(\theta, d)\pi(\theta)d\theta \leqslant \sup_{\theta\in\Omega}\{R(\theta, d)\}.$$

由于 d 的任意性，得到

$$\sup_{\theta\in\Omega}\{R(\theta, d^*)\} = \inf_{d\in\mathscr{D}}\{\sup_{\theta\in\Omega}[R(\theta, d)]\}.$$

由 (7.5.5) 式知，d^* 是参数 θ 的极大极小估计量.

*§7.7 非参数估计

我们对于分布函数 $F(x;\theta)$ 中参数 θ 的估计问题，就其非随机性的与随机性的两方面分别作了讨论. 现在，将讨论另一类型的估计问题，即非参数性的估计问题. 这一节主要叙述由一个随机变数估计或预测另一随机变数的问题. 由一组多个随机变数来估计或预测另一随机变数的问题，其基本原理与上述问题是一致的，只是具体过程较为繁琐，对此我们不作详细讨论. 这节内容，在数理统计学的教科书中，有的称之为多元相关分析，有的称之为第一类回归问题. 我们考虑到这类问题的实际意义在于处理估计与预测问题，因而冠以非参数估计的名称，安排在这一章中作介绍.

一、最小均方误差估计

我们知道，不论随机变数 ξ 及 η 服从什么分布，ξ 及 η 之间下述线性关系式

$$\eta = \alpha\xi + \beta \tag{7.7.1}$$

几乎处处成立的充要条件是其相关系数 ρ 满足条件：$|\rho| = 1$. 其中，α 及 β 为常数. 这个线性关系式的实际意义在于：在 ξ 及 η 中，可由一个随机变数通过它们之间的线性关系来估计或预测另一个

随机变数．这样一种统计模型，称之为线性模型，因而它是线性模型的估计或预测问题．

当$|\rho|<1$时，ξ及η之间的线性关系式不成立．试问，在什么准则下，建立一个随机变数来估计或预测另一个随机变数为"最佳"？例如，在防空中需要随时估计敌机飞行的状态，一般可以将某个时刻的状态表为随机向量$\eta=(\eta_1,\cdots,\eta_r)$，对随机向量$\eta$进行实际观察的结果用另一随机向量$\xi=(\xi_1,\cdots,\xi_s)$表示，通常是$s<r$．就是说，用观测随机变数（或随机向量）$\xi$建立一个统计量$T(\xi)$来对随机变数（或随机向量）$\eta$作出估计或预测．

定义 7.7.1 （最小均方误差估计） 设$T(\xi)$为η的估计量，对于η的任一估计量$U(\xi)$，如果

$$E[\eta-T(\xi)]^2\leqslant E[\eta-U(\xi)]^2, \tag{7.7.2}$$

则称$T(\xi)$为η的最小均方误差估计量．

在第三章的§3.5中已指出，当$T(\xi)=E(\eta|\xi)$时，对任一$U(\xi)$，不等式

$$E[\eta-E(\eta|\xi)]^2\leqslant E[\eta-U(\xi)]^2 \tag{7.7.3}$$

成立．因此，$T(\xi)=E(\eta|\xi)$为η的最小均方误差估计量．

例 7.7.1 设(ξ,η)服从二维正态分布，其联合密度函数为

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-a_1)^2}{\sigma_1^2}-2\rho\frac{(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}\right]\right\},$$

即ξ服从正态$N(a_1,\sigma_1)$，η服从正态$N(a_2,\sigma_2)$，ρ为其相关系数．由§3.5中的结论知道

$$E(\eta|\xi)=a_2+\rho\frac{\sigma_2}{\sigma_1}(\xi-a_1);$$

$$E(\xi|\eta)=a_1+\rho\frac{\sigma_1}{\sigma_2}(\eta-a_2),$$

因而，η 由 ξ 建立的最小均方误差估计量

$$T(\xi)=\left(a_2-\rho\frac{\sigma_2}{\sigma_1}a_1\right)+\rho\frac{\sigma_2}{\sigma_1}\xi \qquad (7.7.4)$$

为一线性模型. 同理，ξ 由 η 建立的最小均方误差估计量也是一个线性模型：

$$T(\eta)=\left(a_1-\rho\frac{\sigma_1}{\sigma_2}a_2\right)+\rho\frac{\sigma_1}{\sigma_2}\eta. \qquad (7.7.5)$$

二、线性最小均方误差估计

我们知道，当 $|\rho|=1$ 或 (ξ,η) 服从二维正态分布时，都可建立线性模型，利用这个模型由一个随机变数在最小均方误差意义下估计或预测另一随机变数，这在实际工作中是相当方便的. 当然，若 $|\rho|<1$ 或 (ξ,η) 不服从二维正态分布时，仍然建立线性模型，就会增加因模型的不合而带来的误差. 由于线性模型在实际工作中应用方便，所以人们对它还是比较有兴趣，讨论得比较详细.

现在，仅就线性模型的估计问题作如下讨论.

定义7.7.2 （线性最小均方误差估计） 设有线性统计量

$$T(\xi)=\alpha+\beta\xi. \qquad (7.7.6)$$

如果对于任一线性统计量 $U(\xi)=\alpha'+\beta'\xi$，都有

$$E[\eta-(\alpha+\beta\xi)]^2\leqslant E[\eta-(\alpha'+\beta'\xi)]^2, \qquad (7.7.7)$$

则称 $T(\xi)$ 为 η 的线性最小均方误差估计量.

显然

$$E[\eta-E(\eta|\xi)]^2\leqslant E[\eta-(\alpha+\beta\xi)]^2. \qquad (7.7.8)$$

因而，线性最小均方误差估计量的精度（在均方误差的意义下）不如最小均方误差估计量. 但当 (ξ,η) 服从二维正态分布时，这两种估计量是一致的.

试问，在线性模型(7.7.6)式中的系数 α 及 β 满足什么条件，可使得 $T(\xi)$ 为 η 的线性最小均方误差估计量. 由于

$$
\begin{aligned}
&E[\eta-(\alpha'+\beta'\xi)]^2\\
=&E\{[\eta-E(\eta)]-\beta'[\xi-E(\xi)]-[\alpha'-E(\eta)+\beta'E(\xi)]\}^2\\
=&E[\eta-E(\eta)]^2+\beta'^2E[\xi-E(\xi)]^2\\
&+[\alpha'-E(\eta)+\beta'E(\xi)]^2-2\beta'\cdot\operatorname{cov}(\xi,\eta)\\
=&D(\eta)+\beta'^2D(\xi)-2\beta'\operatorname{cov}(\xi,\eta)+[\alpha'-E(\eta)+\beta'E(\xi)]^2,
\end{aligned}
$$

将上式分别对 α' 及 β' 求偏导数，得到：

$$
\begin{cases}
\dfrac{\partial}{\partial\alpha'}E[\eta-(\alpha'+\beta'\xi)]^2=2[\alpha'-E(\eta)+\beta'E(\xi)];\\
\dfrac{\partial}{\partial\beta'}E[\eta-(\alpha'+\beta'\xi)]^2=2\beta'D(\xi)-2\operatorname{cov}(\xi,\eta)\\
\qquad+2[\alpha'-E(\eta)+\beta'E(\xi)]\cdot E(\xi),
\end{cases}
$$

令其偏导数等于零，从而解得：

$$
\left.\begin{aligned}
&\alpha=E(\eta)-\beta E(\xi);\\
&\beta=\frac{\operatorname{cov}(\xi,\eta)}{D(\xi)}=\rho\frac{\sqrt{D(\eta)}}{\sqrt{D(\xi)}}.
\end{aligned}\right\}\tag{7.7.9}
$$

可以验证，(7.7.9)所确定的 α 及 β 使得(7.7.7)式成立，即 $T(\xi)=\alpha+\beta\xi$ 为 η 的线性最小均方误差估计量.

同理，可求出 ξ 的线性最小均方误差估计量：

$$
T_1(\eta)=\alpha_1+\beta_1\eta,\tag{7.7.10}
$$

其中

$$
\begin{aligned}
\alpha_1&=E(\xi)-\beta_1E(\eta),\\
\beta_1&=\frac{\operatorname{cov}(\xi,\eta)}{D(\eta)}=\rho\frac{\sqrt{D(\xi)}}{\sqrt{D(\eta)}}.
\end{aligned}\tag{7.7.11}
$$

从(7.7.4)与(7.7.9)、(7.7.5)与(7.7.11)四式再次看到，当 (ξ,η) 服从二维正态分布时，最小均方误差估计量与线性最小均方误差估计量是一致的.

现在讨论由(7.7.9)确定的线性估计量的性质. 因为

$$
E[\alpha+\beta\xi]=E[E(\eta)-\beta E(\xi)+\beta\xi]
$$

$$=E(\eta)-\beta E(\xi)+\beta E(\xi)=E(\eta),$$

即由(7.7.9)确定的线性统计量 $T(\xi)=\alpha+\beta\xi$ 是 η 的线性最小均方误差无偏估计量，简称为 η 的最优线性估计量. 记

$$e=\eta-(\alpha+\beta\xi),$$

称 e 为用 $T(\xi)=\alpha+\beta\xi$ 作为 η 的最优线性估计量的估计误差，它具有下述性质：

(1) $E(e)=0, D(e)=(1-\rho^2)D(\eta)$; (7.7.12)

(2) $\operatorname{cov}(e,\xi)=0$. (7.7.13)

事实上，$E(e)=0$ 是显然的. $D(e)$ 由下式计算得到：

$$\begin{aligned}D(e)&=E[\eta-(\alpha+\beta\xi)]^2=E\{\eta-[E(\eta)-\beta E(\xi)+\beta\xi]\}^2\\&=E\{[\eta-E(\eta)]-\beta[\xi-E(\xi)]\}^2\\&=D(\eta)-2\beta\operatorname{cov}(\xi,\eta)+\beta^2D(\xi)\\&=D(\eta)-2\rho^2\cdot D(\eta)+\rho^2D(\eta)=(1-\rho^2)\cdot D(\eta).\end{aligned}$$

(7.7.13)式的证明如下：因为

$$e=\eta-(\alpha+\beta\xi)=[\eta-E(\eta)]-[\xi-E(\xi)]\frac{\operatorname{cov}(\xi,\eta)}{D(\xi)},$$

所以

$$\begin{aligned}\operatorname{cov}(e,\xi)&=E[e-E(e)][\xi-E(\xi)]=E\{e[\xi-E(\xi)]\}\\&=E[\eta-E(\eta)][\xi-E(\xi)]-E[\xi-E(\xi)]^2\cdot\frac{\operatorname{cov}(\xi,\eta)}{D(\xi)}\\&=\operatorname{cov}(\xi,\eta)-\operatorname{cov}(\xi,\eta)=0,\end{aligned}$$

所以估计误差 e 与 ξ 线性不相关.

综上所述，用(7.7.9)确定的 η 的最优线性估计 $T(\xi)=\alpha+\beta\xi$，它的估计误差 $e=\eta-T(\xi)$ 的数学期望为零，方差随着 ρ 增大而减少，e 与 ξ 线性不相关.

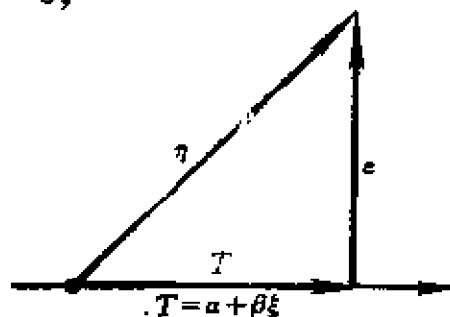


图 7.7.1

借助于几何的观点，把不相关性视为正交性，那么估计误差 e 可视为与 ξ 正交(即垂直)．这样，η 与 ξ 本来不正交，但 $e=\eta-T$ 与 ξ 正交，而 $T=\alpha+\beta\xi$ 为 η 在 ξ 上的垂直投影．就是说，由 ξ 建立的 η 的最优线性估计 $T(\xi)=\alpha+\beta\xi$ 是 η 在 ξ 上的垂直投影．

同理，由(7.7.11)确定的 ξ 的最优线性估计 $T_1(\eta)\overset{\triangle}{=}\alpha_1+\beta_1\eta$，它是 ξ 在 η 上的垂直投影．

习　题

1．设某地区的风速 ξ 服 $\Gamma(\alpha,\beta)$，密度函数为

$$f(x,\alpha,\beta)=\begin{cases}\dfrac{\beta^{-(\alpha+1)}}{\Gamma(\alpha+1)}x^{\alpha}e^{-\frac{x}{\beta}}, & x>0;\\ 0 & ,x\leqslant 0.\end{cases}$$

$(\alpha+1>0,\beta>0)$，试按频率估计概率的原理，在 $\beta=\dfrac{1}{9}$，$\alpha=2$ 时，$n=5$ 求百年一遇的最大风速 M 值(即 ξ 大于 M 的概率为 1%)．

2．对任一地区，地震的震级数 y 与其发生的次数 n 之间有经验公式：$n(y)=e^{a-\beta y}$，$a>0,\beta>0,y\geqslant 0$，试按频率估计概率的原理，求震级 ξ 的分布函数．

3．设总体 ξ 服从正态 $N(a,1)$，今观察了二十次，只记录是否为负值，若事件"$\xi<0$"出现了十四次，试按频率估计概率的原理，试求 a 的估计值．

4．设总体 ξ 的密度函数为 $f(x;\theta)$，$\xi_1,\cdots,\xi_n$ 为其子样，求参数 θ 的极大似然估计量．

(1) $f(x;\theta)=\dfrac{1}{2}e^{-|x-\theta|}$，$|x|>\theta$，$-\infty<\theta<\theta$，$-\infty<x<\infty$；

(2) $f(x;\theta)=\begin{cases}\theta x^{\theta-1}, 0<x<1 & 0<\theta<\infty,\\ 0, & 其它;\end{cases}$

(3) $f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & 0<x\leqslant\theta \quad 0<\theta<\infty,\\ 0, & 其它;\end{cases}$

(4) $f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & 0<x<\infty \quad 0<\theta<\infty,\\ 0, & 其它.\end{cases}$

5. 设总体ξ服从二项分布$b(N,p)$，$0<p<1$，N为正整数，$\xi_1,\cdots,\xi_n$为其子样，求N及p的矩法估计量.

6. 设总体ξ服从对数正态分布，密度函数为

$$f(x;a,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(\lg x-a)^2\right\},$$

$\xi_1,\cdots,\xi_n$为其子样，求a及σ的矩法估计量.

7. 设总体ξ的密度函数为

$$f(x)=(\beta+1)x^\beta,\quad 0<x<1,\quad \beta>-1.$$

$\xi_1,\cdots,\xi_n$为其子样，求参数β的极大似然估计及矩法估计. 今得子样观察值为0.3、0.8、0.27、0.35、0.62及0.55，求参数β的估计值.

8. 设T为电子元件的失效时间(单位:小时)，其密度函数为

$$f(t)=\beta e^{-\beta(t-t_0)},\quad t>t_0>0$$

(即随机变数T具有在左边t_0截头的，参数为β的指数分布). 假定n个元件独立地试验并记录其失效时间分别为$T_1,\cdots,T_n$.

(1) 当t_0为已知时，求β的极大似然法估计量.

(2) 当β为已知时，求t_0的极大似然法估计量.

9. 设总体ξ服从韦布分布，密度函数为:

$$f(x;\lambda)=(\lambda\alpha)x^{\alpha-1}e^{-\lambda x^\alpha},\quad x>0,$$

$\xi_1,\cdots,\xi_n$为其子样. 若α为已知，求参数λ的极大似然法估计量.

10. 设总体ξ服从Γ-分布，密度函数为:

$$f(x;\beta)=\frac{\beta^{-(\alpha+1)}}{\Gamma(\alpha+1)}x^\alpha e^{-\frac{x}{\beta}},\quad x>0,$$

$\xi_1,\cdots,\xi_n$为其子样. 若α为已知，求参数β的极大似然法估计量.

11. 设总体ξ的密度函数为:

$$f(x;\theta_1,\theta_2)=\frac{1}{\theta_2}e^{-\frac{x-\theta_1}{\theta_2}},$$

$-\infty<\theta_1<x<\infty$，$0<\theta_2<\infty$，$\xi_1,\cdots,\xi_n$为其子样，求参数$\theta_1$及$\theta_2$的极大似然法估计量.

12. 设总体ξ服从正态$N(a,\sigma)$，$\xi_1,\cdots,\xi_n$为其子样.

(1) 求k，使$\hat{\sigma}=\frac{1}{k}\sum_{i=1}^{n}|\xi_i-\bar{\xi}|$为$\sigma$的无偏估计量.

(2) 求 k，使 $\hat{\sigma}^2=\frac{1}{k}\sum_{i=1}^{n-1}(\xi_{i+1}-\xi_i)^2$ 为 σ^2 的无偏估计量.

13. 设总体 ξ 的数学期望为 a，方差为 σ^2，ξ_1，…，ξ_n 是它的子样，$T(\xi_1,\cdots,\xi_n)$ 为 a 的任一线性无偏估计量. 证明子样平均 $\bar{\xi}$ 与 T 的相关系数为 $\sqrt{D(\bar{\xi})/D(T)}$.

14. 设总体 ξ 服从正态 $N(a,\sigma)$，$\xi_1,\cdots,\xi_n$ 为其子样，a 为已知，证明 $\frac{1}{n}\sqrt{\frac{\pi}{2}}\sum_{i=1}^{n}|\xi_i-a|$ 为 σ 的无偏估计量，且有效率为 $\frac{1}{\pi-2}$.

15. 设总体服从正态 $N(a,1)$，$-\infty<a<\infty$，ξ_1,ξ_2,ξ_3 为其子样，试证下述三个估计量

(1) $\hat{a}_1=\frac{1}{5}\xi_1+\frac{3}{10}\xi_2+\frac{1}{2}\xi_3$;

(2) $\hat{a}_2=\frac{1}{3}\xi_1+\frac{1}{4}\xi_2+\frac{5}{12}\xi_3$;

(3) $\hat{a}_3=\frac{1}{3}\xi_1+\frac{1}{6}\xi_2+\frac{1}{2}\xi_3$

都是 a 的无偏估计量，并求出每一估计量的方差，问哪一个最小？

16. 设总体 ξ 的数学期望为 a，$\hat{a}_1$ 及 $\hat{a}_2$ 分别为参数 a 的两个无偏估计量，它们的方差分别为 σ_1^2 及 σ_2^2，相关系数为 ρ，试确定常数 $c_1>0$，$c_2>0$，$c_1+c_2=1$，使得 $c_1\hat{a}_1+c_2\hat{a}_2$ 有最小方差.

17. 设总体 ξ 服从正态 $N(a_1,1)$，总体 η 服从正态 $N(a_2,2)$，$\xi_1,\cdots,\xi_{n_1}$ 为总体 ξ 的子样，$\eta_1,\cdots,\eta_{n_2}$ 为总体 η 的子样，且这两个子样相互独立.

(1) 试求 $a=a_1-a_2$ 的无偏估计量 $\hat{a}$.

(2) 如果 $n_1+n_2=n$ 固定，问 n_1 及 n_2 如何配置，可使 $\hat{a}$ 的方差达到最小.

18. 设总体 ξ 及 η 的数学期望及方差分别为 a_1，a_2 及 σ_1^2，σ_2^2，$\xi_1,\cdots,\xi_n$ 及 $\eta_1,\cdots,\eta_n$ 分别为它们的子样，这两个子样相互独立. $\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$，$\bar{\eta}=\frac{1}{n}\sum_{i=1}^{n}\eta_i$，试证 $\bar{\xi}-\bar{\eta}$ 为 a_1-a_2 的最优线性无偏估计量.

19. 设总体 ξ 的密度函数为:

$$f(x;\theta)=\begin{cases}1, & \theta-\frac{1}{2}\leqslant x\leqslant\theta+\frac{1}{2};\\0, & 其它,\end{cases}$$

$-\infty<\theta<\infty$, $\xi_1,\cdots,\xi_n$ 为其子样.

(1) 求参数 θ 的极大似然估计量.

(2) 证明子样平均 $\bar{\xi}$ 及 $\frac{1}{2}[\max\limits_{1\leqslant i\leqslant n}\xi_i+\min\limits_{1\leqslant i\leqslant n}\xi_i]$ 都是 θ 的无偏估计量，问哪个较有效?

20. 设总体 ξ 的密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}, & 0<x\leqslant\theta, 0<\theta<\infty;\\0, & 其它,\end{cases}$$

ξ_1、ξ_2、ξ_3 为其子样. 试证 $\frac{4}{3}\max\limits_{1\leqslant i\leqslant 3}\xi_i$ 及 $4\min\limits_{1\leqslant i\leqslant 3}\xi_i$ 都是参数 θ 的无偏估计量，问哪个较有效?

21. 设 $\hat{\theta}_1$ 及 $\hat{\theta}_2$ 是参数 θ 的两个独立的无偏估计量，且 $\hat{\theta}_1$ 的方差为 $\hat{\theta}_2$ 的方差的两倍，试确定常数 c_1 及 c_2，使得 $c_1\hat{\theta}_1+c_2\hat{\theta}_2$ 为参数 θ 的无偏估计量，并且在所有这样的线性估计中方差最小.

22. 试证 (7.2.18) 式成立.

23. 设 T_1 及 T_2 分别是参数 θ 的可估计函数 $g_1(\theta)$ 及 $g_2(\theta)$ 的最优无偏估计量，试证 $b_1T_1+b_2T_2$ 是 $b_1g_1(\theta)+b_2g_2(\theta)$ 的最优无偏估计量，其中 b_1 和 b_2 是常数.

24. 试证 (7.2.19) 及 (7.2.20) 两式成立.

25. 设总体 ξ 服从正态 $N(a_1,\sigma_1)$，总体 η 服从正态 $N(a_2,\sigma_2)$，$\xi_1,\cdots,\xi_{n_1}$ 及 $\eta_1,\cdots,\eta_{n_2}$ 分别为其子样，并且这两个子样相互独立.

(1) 试建立 σ_1^2/σ_2^2 的置信水平为 $1-\alpha$ 的区间估计;

(2) 假定 $\sigma_1=\sigma_2$，试建立 a_1-a_2 的置信水平为 $1-\alpha$ 的区间估计.

26. 设总体 ξ 服从正态 $N(a_1,\sigma)$，总体 η 服从正态 $N(a_2,\sigma)$，其中 σ 未知，$\xi_1,\cdots,\xi_n$ 及 $\eta_1,\cdots,\eta_n$ 分别为其子样，并且这两个子样相互独立，求 a_1-a_2 的置信水平为 0.95 的区间估计.

27. 设总体 ξ 服从正态 $N(a,\sigma)$，已知 $\sum\limits_{i=1}^{15}x_i=8.7$，$\sum\limits_{i=1}^{15}x_i^2=25.05$，试分别求置信水平为 0.95 的 a 及 σ^2 的区间估计.

28. 若从自动车床加工的一批零件中随机抽取10个，测得其尺寸与规定尺寸的偏差(单位：微米)分别为：2、1、-2、3、2、4、-2、5、3、4. 零件尺寸的偏差记作ξ，假设总体ξ服从正态$N(a,\sigma)$，试求a及σ^2的无偏估计值，并求置信水平为0.9的区间估计.

29. 设总体ξ服从指数分布，即

$$f(x)=\lambda e^{-\lambda x},\quad x>0,\quad \lambda>0,$$

损失函数为二次式，λ的先验分布为$\Gamma(\alpha,\beta)$，即

$$g(\lambda)=\frac{\beta^{-(\alpha+1)}}{\Gamma(\alpha+1)}\lambda^{\alpha}e^{-\frac{\lambda}{\beta}},\quad \lambda>0,\beta>0,\alpha+1>0.$$

试求参数λ的贝叶斯估计.

30. 设总体ξ服从正态$N(a,1)$，$-\infty<a<\infty$，ξ_1,ξ_2为其子样，考虑下述三个估计量

$$\hat{a}_1=d_1(\xi_1,\xi_2)=\frac{2}{5}\xi_1+\frac{3}{5}\xi_2;$$

$$\hat{a}_2=d_2(\xi_1,\xi_2)=\frac{5}{6}\xi_1+\frac{1}{6}\xi_2;$$

$$\hat{a}_3=d_3(\xi_1,\xi_2)=\frac{1}{3}\xi_1+\frac{2}{3}\xi_2.$$

定义损失函数为

$$L(a,\hat{a})=3a^2(a-\hat{a})^2,$$

试求$R(d_1,a)$、$R(d_2,a)$、$R(d_3,a)$，问哪个风险为最小？

第八章　假设检验

§8.1　引　　言

假设检验同估计理论一样，都是数理统计学的重要内容．设总体 ξ 的分布函数 $F(x;\theta)$ 中 θ 为未知参数，现对未知参数 θ 提出假设“θ_0 为其真值”，试问怎样利用子样 $\xi_1,\cdots,\xi_n$ 提供的信息来检验这个假设应否定还是不应否定，“最佳”检验的准则又怎样确定？这类统计问题，称之为参数的假设检验问题．参数检验与参数估计，是利用子样 $\xi_1,\cdots,\xi_n$ 对总体的概率统计特性提供的信息，建立子样的函数——估计量或检验统计量，从不同的角度来处理未知参数 θ 的两种统计方法．

除参数的假设检验问题之外，非参数的假设检验问题也相当重要．例如，若总体 ξ 的分布函数表达式 $F(x)$ 不知道时，假设总体 ξ 的分布函数为某个指定的分布函数 $F_0(x)$，问怎样利用子样 $\xi_1,\cdots,\xi_n$ 提供的信息来对假设作出判断，是否定或不否定，就是一个非参数性假设检验．又如随机变数 ξ 与 η 之间的相关性或独立性等问题的假设检验，也归属于非参数性的假设检验．

关于检验法的准则，将引进最佳检验、一致最佳检验及无偏检验等概念．

现在，我们从几个例子谈起．

例 8.1.1 某工厂金工车间生产一种铆钉，铆钉直径的标准定为 $a_0=2$ 厘米．现为了提高产量，采用了一种新工艺，现从新工艺生产的铆钉中抽取 100 个，测其直径计算得平均值 $\bar{x}=1.978$ 厘米．我们要问，$\bar{x}$ 与 a_0 之间的差异，纯粹是试验或测试的误差造成

的，还是反映了工艺条件的改革使得铆钉直径发生了显著性的变化.

上述例子是关于数学期望 $E(\xi)=2$ 的假设检验问题. 凡对随机变数 ξ 的数字特征或分布函数中的参数提出假设作检验的问题，都称之为参数性的假设检验问题. 又如下面的问题，是非参数性的假设检验问题，它是对随机变数 ξ 的分布函数表达式提出假设作检验的问题.

例 8.1.2 认为某工厂生产的灯泡其光通量 ξ 服从正态分布，是否正确，如何检验；又认为某电话交换台在某段时间接到的呼唤次数 η 服从泊松分布，是否正确，如何检验？

关于假设检验的思想方法，没有参数估计那样直观. 现在，我们着重讨论例 8.1.1，阐述假设检验这种统计方法的基本思想及主要步骤，从而归纳出数学模型.

(1) 假设检验的基本思想

在例 8.1.1 中，从新工艺生产的铆钉来看，$\bar{x}$ 比 a_0 小了些，其差为 $a_0-\bar{x}=0.022$ 厘米. 前面提出这个差异是新工艺造成的还是纯粹由于随机因素引起的问题，可以用"假设检验"这种统计方法对此作出判断. 假设检验的基本思想是这样：用 a_0 表示原工艺生产的铆钉直径总体 ξ 的数学期望 $E(\xi)$，假设"新工艺对铆钉直径没有显著影响"，那么从新工艺生产的铆钉中抽取的子样，则可认为是从原工艺总体 ξ 中抽取的. 假定总体 ξ 服从正态 $N(a_0,\sigma_0)$，其中 σ_0 为已知(如：由以往生产情况确定)，那么统计量

$$U=\frac{\bar{\xi}-a_0}{\sigma_0/\sqrt{n}} \tag{8.1.1}$$

服从正态 $N(0,1)$. 由于 $P\{|U|\leqslant 1.96\}=0.95$，所以

$$P\left\{a_0-1.96\frac{\sigma_0}{\sqrt{n}}\leqslant\bar{\xi}\leqslant a_0+1.96\frac{\sigma_0}{\sqrt{n}}\right\}=0.95. \tag{8.1.2}$$

因此，若"假设"是正确的，那么从新工艺生产的铆钉总体中抽取容量为 n 的子样 $\xi_1, \cdots, \xi_n$，其子样平均 $\bar{\xi}$ 的观察值 $\bar{x}$ 落在这个区间之外的概率为 0.05. 对此可作如下解释："在抽取子样 20 次中，大约只有 1 次 $\bar{x}$ 值落在这个区间之外". 如果在一次实际抽取的容量为 n 的子样的平均值 $\bar{x}$ 落在这个区间外面，那么我们就说从新工艺总体中抽取的子样不能"认为"来自原工艺生产的铆钉总体 ξ，即认为"假设"是不正确的，则"假设"应该被否定或者说被拒绝.

(2) 判断"假设"的根据

上述判断假设是否定或不否定，是根据所谓小概率事件的原理. 小概率事件原理是："小概率事件(或概率很小的事件)在一次试验(或观察)中是几乎不可能发生的". 我们再详细解释一下，设有某假设 H_0 需要检验，先假设 H_0 是正确的，在此"假设"之下，构造某一个事件 A，它在 H_0 为正确的条件下的概率很小，例如 $P\{A|H_0\}=0.05$，现在进行一次试验，如果事件 A 发生了，就说一个小概率事件居然发生了，这就不能不使人怀疑 H_0 的正确性，因而很可能要否定 H_0.

下面，再举一例予以说明.

例 8.1.3 某箱子中有白球及黑球，总数为 100，但不知白球及黑球各占多少. 现在提出假设 H_0："其中 99 个是白的". 下面用刚才的想法来判断这个假设 H_0 是否正确. 暂设 H_0 正确，那么从箱子中任取一球，"得黑球"的概率为 0.01，我们认为这是一小概率事件. 如果抽一球居然抽得是黑球，那么自然使人们要否定 H_0，就是说白球的个数不是 99.

(3) 第一类错误

按上述原则否定 H_0，不免要犯错误. 我们知道，当 H_0 正确时，小概率事件 A 也有可能发生，并非 A 绝对不发生. 因此，当 H_0 本来是正确的，而在事件 A 发生后，却错误地否定了 H_0，这种"弃

真"的错误，称之为犯第一类错误．由上所述，犯第一类错误的概率为 α，即 $P\{A|H_0\}=\alpha$．例如在检验新工艺的例 8.1.1 中，由(8.1.2)式知道，犯第一类错误的概率为 $\alpha=0.05$．

现将上述讨论，归纳出假设检验问题的三个主要步骤：

第一步 提出原假设

$$H_0: a=a_0 \quad 或 \quad H_0: F(x)=F_0(x)$$

如例 8.1.1 中，$H_0: a=2$．例 8.1.2 中，$H_0: F(x)=F_0(x)$，对于光通量 ξ 来说，$F_0(x)$ 是正态分布，对于呼唤次数 η 来说，$F_0(x)$ 是泊松分布．可见，原假设（又称作零假设）H_0 是我们所要进行检验的对象．

第二步 建立检验统计量

建立检验统计量是假设检验中重要的环节．比如例 8.1.1 中，在总体 ξ 服从正态 $N(a,\sigma_0)$ 的假定下，当原假设 $H_0: a=a_0$ 成立时，建立检验统计量 $U=\dfrac{\bar{\xi}-a_0}{\sigma_0/\sqrt{n}}$，$U$ 服从标准正态 $N(0,1)$．注意，检验统计量是子样的函数，要求不带有任何未知参数．

对于总体 ξ 的分布函数 $F(x;\theta_1,\theta_2)$ 中参数 θ_1、θ_2 的假设检验，在 ξ 的分布函数为正态 $N(a,\sigma)$ 的基本假定下，常用的检验统计量有 t-分布、χ^2 分布、F-分布，这些适合于小样问题．如果总 ξ 不服从正态分布，或总体 ξ 的分布函数未知，这时检验统计量的精确分布难于求出或相当复杂，如有可能求出其渐近分布，则只适用于大样问题．非参数性的检验问题，一般都是大样问题，如 § 8.3 中所讨论的检验问题．

第三步 确定 H_0 的否定域

如例 8.1.1 中，当原假设 H_0 成立时，检验统计量 U 服从正态 $N(0,1)$，那么给定满足 $0<\alpha<1$ 的 α 值，在标准正态分布表中查得临界值 u_α，使得

$$P\{|U|\geq u_\alpha\}=\alpha,$$

或者

$$P\{-u_\alpha<U\leq u_\alpha\}=1-\alpha. \tag{3.1.3}$$

若由子样 $\xi_1,\cdots,\xi_n$ 的观察值 $x_1,\cdots,x_n$，算得统计量U的值 u 落在$(-\infty,-u_\alpha)$或(u_α,∞)时则否定 H_0. 称$(-\infty,-u_\alpha)$及(u_α,∞)组成为 H_0 的否定域，称 u_α 为临界值.

对原假设 H_0 作出否定或不否定的判断，通常称之为对 H_0 作显著性检验. 称 α 为显著性水平，1-α 为置信水平. 以后常用"在显著性水平 α 下对 H_0 作显著性检验"这类术语. 值得强调的是，**我们对 H_0 作出判断，是冒着犯第一类错误的风险的.**

对于 H_0，给定不同的显著性水平 α，对应有不同的临界值 u_α，相应地有不同的否定域，因而有不同的判断结论，这点必须注意，如下图所示

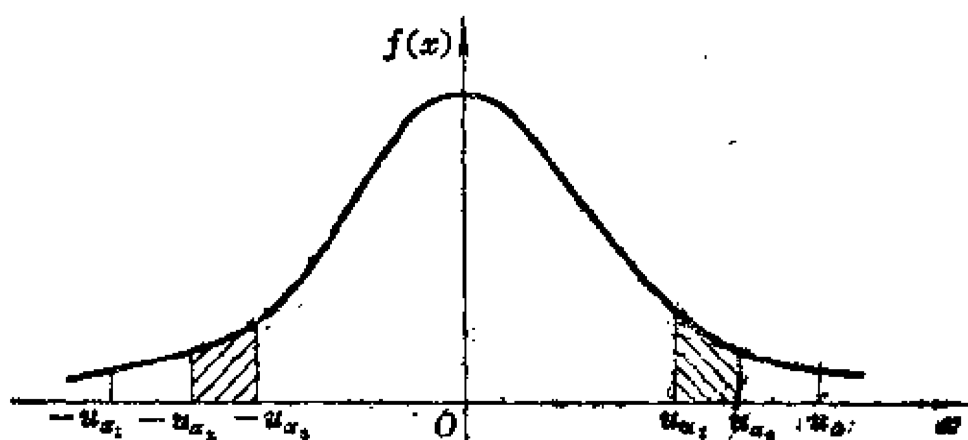


图 3.1.1

$$\alpha_1=0.01,\quad u_{\alpha_1}=2.58$$
$$\alpha_2=0.05,\quad u_{\alpha_2}=1.96$$
$$\alpha_3=0.10,\quad u_{\alpha_3}=1.65$$

如果检验统计量U的值u落在阴影部分，在给定显著性水平$\alpha_3=0.10$时，作出否定 H_0 的结论；而在给定显著性水平 $\alpha_2=0.05$ 时，作出不否定 H_0 的结论.

还必须指出，不否定 H_0，绝不意味着不必再作进一步的考察就可以断然接受 H_0 这个假设，即并不认为 H_0 一定是正确的. 这个观点必须注意，我们将在§3.4中作较深入的讨论. 在此，用图 3.1.2 作简略的说明.

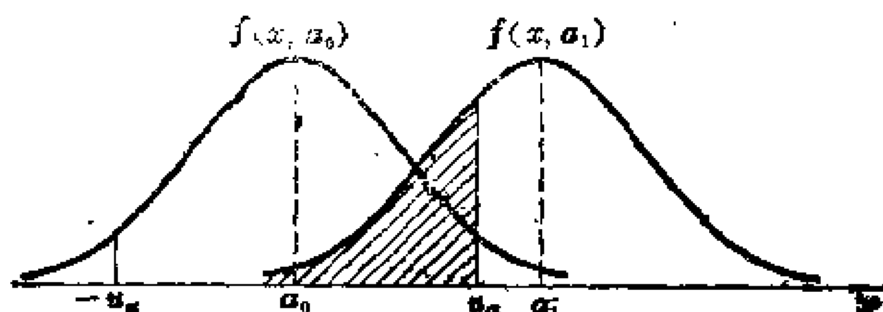


图 8.1.2

如果检验统计量U的值$u\in(-u_\alpha,u_\alpha)$，则不否定H_0，但还不能认为H_0是正确的. 因为当u落在阴影部分时，子样平均值$\bar{\xi}$来自服从正态$N(a_0,\sigma)$的总体ξ的可能性很大，但来自服从正态$N(a_1,\sigma)$的可能性也很大，也就是说，总体的数学期望$E(\xi)=a$可能是取a_0值，但不能排斥取a_1值的可能性.

§8.2 参数假设检验

参数假设检验的一般提法是：设总体ξ的分布函数$F(x;\theta)$中，θ为未知参数，$\theta\in\Omega$，Ω为参数空间. 我们将参数空间Ω分解为互不相交的两个部分Ω_0及$\Omega-\Omega_0$，即$\Omega_0\cap(\Omega-\Omega_0)=\varnothing$，$\Omega_0\cup(\Omega-\Omega_0)=\Omega$. 考虑检验问题：

$$H_0:\theta\in\Omega_0,\qquad H_1:\theta\in\Omega-\Omega_0,$$

Ω_0为非空子集，Ω_0是假设检验的对象，称H_0为原假设(或零假设)，称H_1为备选假设.

如果Ω只含有两个点，即若$\Omega=\{\theta_0,\theta_1\}$，$\Omega_0=\{\theta_0\}$，$\Omega-\Omega_0=\{\theta_1\}$，则有

$$H_0:\theta=\theta_0,\qquad H_1:\theta=\theta_1.$$

这时称H_0及H_1分别为简单原假设及简单备选假设.

如果Ω多于两个点，$\Omega_0=\{\theta_0\}$，而$\Omega-\Omega_0=\{\theta\neq\theta_0\}$为非单点集，即有：

$$H_0:\theta=\theta_0,\qquad H_1:\theta\neq\theta_0,$$

则称 H_0 为简单原假设，H_1 为复合备选假设。

若 Ω_0 及 $\Omega-\Omega_0$ 都是非单点集，则称 H_0 及 H_1 都是复合的。

一、数学期望 a 的检验问题

下面对常用的检验问题加以介绍。为了简便叙述，对每一种检验问题，着重给出检验统计量，并举例说明其应用。

第一种情形，一个总体时 a 的检验：

$$H_0: a=a_0, \qquad H_1: a\neq a_0.$$

(1) 设方差 $D(\xi)=\sigma^2$ 为已知，记 $\sigma^2=\sigma_0^2$

假定总体 ξ 服从正态 $N(a,\sigma_0)$，那么由于子样 $\bar{\xi}$ 服从正态 $N\left(a,\frac{\sigma_0}{\sqrt{n}}\right)$，所以

$$U=\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}}$$

服从正态 $N(0,1)$。因此，当原假设 H_0 成立时，统计量

$$U=\frac{\bar{\xi}-a_0}{\sigma_0/\sqrt{n}} \tag{8.2.1}$$

服从正态 $N(0,1)$，U 不带有未知参数，它可作为判断 H_0 的检验统计量。这种检验法称之为 u 检验法。

(2) 设方差 $D(\xi)=\sigma^2$ 为未知

假定总体 ξ 服从正态 $N(a,\sigma)$，由定理 6.3.1 的推论 1 知

$$\sqrt{n-1}\frac{\bar{\xi}-a}{S}\sim t_{(n-1)}.$$

因此，当原假设 H_0 成立时，记

$$T=\sqrt{n-1}\frac{\bar{\xi}-a_0}{S}\sim t_{(n-1)}, \tag{8.2.2}$$

即统计量 T 服从自由度为 $n-1$ 的 t-分布，且不带有未知参数，它可作为判断 H_0 的检验统计量，这种检验法，称之为 t 检验法。

例 8.2.1 某工厂生产一批产品，质量检查规定：其次品率 p_0

$\leqslant 0.05$，则这批产品可出厂，否则不能出厂. 现从这批产品中抽查50件产品，发现有4件产品是次品，试问这批产品能否出厂. 我们将此问题归结为如下的参数检验问题：

$$H_0: p_0 \leqslant 0.05, \qquad H_1: p_0 \geqslant 0.05. \tag{8.2.3}$$

这个问题将放在§8.4中讨论.

例 8.2.2 某工厂生产的电灯泡的使用时数用 ξ 表示，假定 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是未知参数，现在观察 $n=20$ 个灯泡，测得 20 个使用时数 $x_1,\cdots,x_{20}$，并由此算得 $\bar{x}=1832$，$S=497$，试问该厂电灯泡的平均使用时数为"$a=2000$(小时)"这个结论是否成立？

解 考虑检验问题如下：

$$H_0: a=2000, \qquad H_1: a\neq 2000.$$

由(8.2.2)式计算得 $t_{19}=\sqrt{19}\dfrac{1832-2000}{497}=-1.473$，给定显著性水平 $\alpha=0.05$，对于自由度19，在 t 分布临界值表中，查得临界值 $t_{19}(0.05)=2.093$. 由于 $|t_{19}|=1.473<2.093$，因而不否定 H_0.

第二种情形，两个总体时 a 的比较：

$$H_0: a_1=a_2, \qquad H_1: a_1\neq a_2.$$

现在讨论两个总体的数学期望 a_1 及 a_2 是否相等的假设检验. 假定总体 ξ 服从正态 $N(a_1,\sigma_1)$，总体 η 服从正态 $N(a_2,\sigma_2)$，其中 a_1、σ_1 及 a_2、σ_2 都是未知的. 现在独立地分别从总体 ξ 及 η 中抽取子样 $\xi_1,\cdots,\xi_{n_1}$ 及 $\eta_1,\cdots,\eta_{n_2}$，记

$$\bar{\xi}=\frac{1}{n_1}\sum_{i=1}^{n_1}\xi_i, \quad S_1^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(\xi_i-\bar{\xi})^2;$$

$$\bar{\eta}=\frac{1}{n_2}\sum_{i=1}^{n_2}\eta_i, \quad S_2^2=\frac{1}{n_2}\sum_{i=1}^{n_2}(\eta_i-\bar{\eta})^2.$$

下面，我们分三种情形来讨论 H_0 的显著性检验.

(1) 设 $\sigma_1^2=\sigma_2^2=\sigma^2$.

由(6.3.10)式知，统计量

$$\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sqrt{n_1S_1^2+n_2S_2^2}}\sim t_{(n_1+n_2-2)}. \quad (8.2.4)$$

因此，当 H_0 成立时，统计量

$$\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\frac{(\bar{\xi}-\bar{\eta})}{\sqrt{n_1S_1^2+n_2S_2^2}}\sim t_{(n_1+n_2-2)}. \quad (8.2.5)$$

它不带有未知参数，因而可作为 H_0 的检验统计量，这是 t 检验法.

当 $n_1=n_2=n$ 时，则有 $\sqrt{n-1}\frac{\bar{\xi}-\bar{\eta}}{\sqrt{S_1^2+S_2^2}}\sim t_{2n-2}$.

(2) 设 $\sigma_1^2\neq\sigma_2^2$, $n_1=n_2=n$.

我们采用所谓配对试验的 t 检验法. 定义

$$Z_i=\xi_i-\eta_i\ (i=1,\cdots,n), \quad (8.2.6)$$

记 $E(Z_i)=E(\xi_i-\eta_i)=a_1-a_2=d$, $D(Z_i)=D(\xi_i)+D(\eta_i)=\sigma_1^2+\sigma_2^2=\sigma^2$. 则 $Z_1,Z_2,\cdots,Z_n$ 为总体 Z 服从正态 $N(d,\sigma)$ 的子样. 此时，a_1 与 a_2 是否相等的检验，就等价于下述假设检验：

$$H_0: d=0, \qquad H_1: d\neq 0.$$

因 σ_1 及 σ_2 都未知，则 σ 也未知，这时用 t 检验法. 记

$$\bar{Z}=\frac{1}{n}\sum_{i=1}^{n}Z_i, \qquad S^2=\frac{1}{n}\sum_{i=1}^{n}(Z_i-\bar{Z})^2.$$

当 H_0 成立时，用(8.2.2)式建立的统计量

$$\sqrt{n-1}\frac{\bar{Z}}{S}\sim t_{(n-1)} \quad (8.2.7)$$

作为对 H_0 的检验统计量. (8.2.7)式也可写成为

$$\sqrt{n-1}\frac{\bar{\xi}-\bar{\eta}}{\sqrt{S_1^2+S_2^2-2S_{12}}}\sim t_{n-1},\text{ 其中}$$

$$S_{12}=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\eta_i-\bar{\eta})$$

(3) 设 $\sigma_1^2 \neq \sigma_2^2$, $n_1 \neq n_2$.

下面是斯切非(Scheffé)的解法. 不妨假定 $n_1 < n_2$, 定义:

$$Z_i = \xi_i - \sqrt{\frac{n_1}{n_2}}\eta_i + \frac{1}{\sqrt{n_1 \cdot n_2}} \sum_{k=1}^{n_1} \eta_k - \frac{1}{n_2}\sum_{k=1}^{n_2}\eta_k \quad (i = 1, \cdots, n_1), \tag{8.2.8}$$

则有

$$E(Z_i) = a_1 - \sqrt{\frac{n_1}{n_2}}a_2 + \sqrt{\frac{n_1}{n_2}}a_2 - a_2 = a_1 - a_2 = d,$$

$$D(Z_i) = E\left[\xi_i - a_1 - \sqrt{\frac{n_1}{n_2}}(\eta_i - a_2) + \frac{1}{\sqrt{n_1 \cdot n_2}} \sum_{k=1}^{n_1}(\eta_k - a_2) - \frac{1}{n_2}\sum_{k=1}^{n_2}(\eta_k - a_2)\right]^2$$

$$= \sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2 + \sigma_2^2\left[\frac{n_1}{n_1 n_2} + \frac{n_2}{n_2^2} - \frac{2}{n_2} + \frac{2\sqrt{n_1}}{n_2\sqrt{n_2}} - \frac{2n_1}{n_2\sqrt{n_1 n_2}}\right]$$

$$= \sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2,$$

$$\mathrm{cov}(Z_i, Z_j) = 0 \quad (i \neq j, i, j = 1, 2, \cdots, n_1).$$

于是 $Z_1, \cdots, Z_{n_1}$ 为来自总体 Z 服从正态 $N\left(d, \sqrt{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2}\right)$的子样. 此时,关于数学期望是否相等的假设检验,就等价于考虑下述检验问题:

$$H_0: d = 0, H_1: d \neq 0.$$

记 $\sigma = \sqrt{\sigma_1^2 + \frac{n_1}{n_2}\sigma_2^2}$, σ 未知,这时可用 t 检验法.

记

$$\bar{Z} = \frac{1}{n_1}\sum_{i=1}^{n_1} Z_i, \qquad S^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(Z_i - \bar{Z})^2.$$

则当 H_0 成立时，可用(8.2.2)式建立的统计量

$$\sqrt{n_1-1}\frac{Z}{S}\sim t_{(n_1-1)} \qquad (8.2.9)$$

作为对 H_0 的检验统计量.

例 8.2.3 设有甲、乙两种安眠药，考虑比较它们的治疗效果. 以 ξ 表示失眠病者服甲药后睡眠时间延长的时数；以 η 表示服乙药后睡眠时间延长的时数. 现在独立观察 20 个病者，其中 10 人服甲药，另 10 人服乙药，延长的时数记在下表中：

表 8.2.1

时数 人数 药物	1	2	3	4	5	6	7	8	9	10
x	1.9	0.8	1.1	0.1	0.1	4.4	5.5	1.6	4.6	3.4
y	0.7	−1.6	−0.2	−1.2	−0.1	3.4	3.7	0.8	0.0	2.0

假定 ξ 及 η 分别服从正态 $N(a_1,\sigma)$ 及 $N(a_2,\sigma)$，其中方差 σ^2 相同，试问这两种药物的疗效有无显著性的差异.

解 将上述问题归结为考虑如下检验问题：

$$H_0: a_1=a_2, \qquad H_1: a_1\neq a_2.$$

这可用(8.2.5)式建立的统计量 $t_{(n_1+n_2-2)}$ 对 H_0 作显著性检验.

实际计算结果如下：

$$\bar{x}=2.35, \quad S_1^2=3.52, \quad n_1=10;$$

$$\bar{y}=0.75, \quad S_2^2=2.88, \quad n_2=10.$$

将 $\bar{x}$、S_1^2 及 $\bar{y}$、S_2^2 各值代入(8.2.5)式计算得 $t_{18}=1.90$. 给定显著性水平 $\alpha=0.05$，对于自由度为 18，在 t-分布临界值表中查得临界值 $t_{18}(0.05)=2.10$. 由于 $1.90<2.10$，因而不否定 H_0. 在实际工作中，就认为这两种药物的疗效没有显著性的差异.

例 8.2.4 某工厂生产某种电器材料，要检验原标准材料的平均疲劳寿命同新试验材料的平均疲劳寿命之间有无显著性的差

异. 试验所得的数据列在表 8.2.2 中:

表 8.2.2

标准材料的疲劳寿命(小时)	40	110	150	65	90	210	270
新试验材料的疲劳寿命(小时)	60	150	220	310	380	350	250
	450	110	175				

用 ξ 及 η 分别表示标准材料及新试验材料的疲劳寿命, 一般情况下, 可假定 ξ 及 η 分别服从对数正态分布, 即有 $X=\log\xi$, $Y=\log\eta$, 其中 X 及 Y 分别服从正态 $N(a_1,\sigma)$ 及正态 $N(a_2,\sigma)$ (在此又假定了 X 及 Y 的方差相同).

解 将上述问题归结为考虑下述检验问题:

$$H_0: E(X)=E(Y), \qquad H_1: E(X)\neq E(Y).$$

可用(8.2.5)式建立的统计量 $t_{(n_1+n_2-2)}$ 对 H_0 作显著性检验. 现将 X 及 Y 的试验值(试验结果的数据)列在表 8.2.3 中

表 8.2.3

$x=\log\xi$	1.602	2.141	2.176	1.813	1.954	2.312	2.431
$y=\log\eta$	1.778	2.176	2.342	2.491	2.580	2.544	2.398
	2.653	2.041	2.243				

实际计算结果如下:

$$\bar{x}=2.0484, \quad S_1^2=\frac{0.501}{7}\approx 0.072, n_1=7;$$

$$\bar{y}=2.3246, \quad S_2^2=\frac{0.663}{10}=0.0663, n_2=10;$$

$$t_{(7+10-2)}\approx -2.01.$$

给定显著性水平 $\alpha=0.05$, 对于自由度 15, 在 t-分布临界值表中查得临界值 $t_{15}(0.05)=2.73$. 由于 $|t_{15}|=2.01<2.73$, 因而不否

定 H_0. 在实际工作中,就说还不能认为这两种材料的疲劳寿命有显著性的差异.

例 8.2.5 现要比较甲、乙两种橡胶制成的轮胎的耐磨性.今从甲、乙两种轮胎中各随机抽取八个，各取一个组成一对,再随机选取八架飞机，将八对轮胎随机地搭配给八架飞机，作耐磨性试验,经过了一定时间的起落后，测得轮胎磨损量(单位:毫克)数据如下:

表 8.2.4

甲	4900	5220	5500	6020	6340	7660	8650	4870
乙	4930	4900	5140	5700	6110	6880	7930	5010

试问这两种轮胎的耐磨性有无显著性的差异?

解 用 ξ 及 η 分别表示甲、乙两种轮胎的磨损量，假定 ξ 及 η 分别服从正态 $N(a_1,\sigma)$ 及正态 $N(a_2,\sigma)$，其中方差假定为相同，提出假设检验问题:

$$H_0: a_1=a_2, \qquad H_1: a_1\neq a_2.$$

下面分别两种情形讨论 H_0 的显著性检验.

(1) 试验数据配对分析的结论

记 $Z=\xi-\eta$，用(8.2.7)式建立的统计量 $t_{(n-1)}$作 H_0 的检验统计量，将表 8.2.4 中的数据对应地相减，将 Z 的数值列在表 8.2.5 中:

表 8.2.5

Z	−30	320	360	320	230	780	720	−140

假设检验问题化为

$$H_0: d=0, \qquad H_1: d\neq 0.$$

实际计算结果如下：

$$\bar{Z}=320,\ S^2=89425,$$

$$t_7=\sqrt{7}\frac{\bar{Z}}{S}\approx 2.83.$$

给定显著性水平 $\alpha=0.05$，对于自由度 $8-1=7$，在 t 分布临界值表中查得临界值 $t_7(0.05)=2.365$，由于 $2.83>2.365$，因而否定 H_0，即认为这两种轮胎的耐磨性有显著差异，事实上从表 8.2.4 看到乙轮胎明显地较耐磨.

（2）试验数据不配对分析的结论

所谓不配对作分析，是指将两种轮胎的数据看作来自两个总体的子样，因而用(8.2.5)式建立的统计量 $t_{(n_1+n_2-2)}$ 作 H_0 的检验统计量.

实际计算结果如下：

$$\bar{x}=6145,\qquad S_1^2=1633900;$$

$$\bar{y}=5825,\qquad S_2^2=1053875;$$

$$t_{14}=\frac{320}{619.7}\approx 0.516.$$

给定显著性水平 $\alpha=0.05$，对于自由度为 14，在 t 分布临界值表中查得临界值为 $t_{14}(0.05)=2.145$．由于 $0.516<2.145$，因而不否定 H_0，即认为这两种轮胎的耐磨性差异不显著.

上述两种情形，在同一显著性水平 $\alpha=0.05$ 下所得的结论不一致.当数据配成对子作分析时，自由度减少了，临界值提高了，增加了否定 H_0 的可能性，这是因为在每架飞机上突出了两种轮胎的耐磨性的差异，消除了飞机之间对数据分析的干扰，只要两种轮胎的耐磨性有一定差异，就有可能否定 H_0．而当数据不配成对子作分析时，自由度增加，临界值降低，减少了否定 H_0 的可能性.这时因为不配对时，飞机之间与轮胎之间对数据的影响交织在一

起，不易分辨轮胎之间耐磨性的差异．如果两个子样之间的平均值差异不大，但数据配对后，每个对子内的两个数据差异很大，用配对的检验法，就有可能作出否定 H_0 的结论．如果两个子样之间的平均值差异较大，经配对后每个对子内的两个数据差异仍较大，但若用不配对的检验法，也有可能作出不否定 H_0 的结论．总之，什么情况下用配对的检验法，什么时候用不配对的检验法，要按处理问题的具体要求而定．

两个总体的数学期望是否相等的显著性检验，其实际意义还在于它是一种选优的统计方法．若 H_0 被否定，数学期望之间的差异是显著的，这时就有比较、选优的意义．至于多个总体的数学期望之间差异性的显著性检验，是建立 F 统计量作检验统计量，用所谓分差分析法进行检验，将在第九章的§9.4中作介绍．

二、方差 σ^2 的检验问题

第一种情形：一个总体时 σ^2 的检验：

$$H_0:\ \sigma^2=\sigma_0^2,\qquad H_1:\ \sigma^2\neq\sigma_0^2.$$

假定总体 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是未知的，由(6.3.5)式知

$$\frac{nS^2}{\sigma^2}\sim\chi^2{}_{(n-1)}.$$

因此，当 H_0 成立时，统计量

$$\frac{nS^2}{\sigma_0^2}\sim\chi^2{}_{(n-1)},\tag{8.2.19}$$

即 $\frac{nS^2}{\sigma_0^2}$ 服从自由度为 $n-1$ 的 χ^2-分布，且不带有未知参数，因此可作为 H_0 的检验统计量，用它来进行显著性检验的方法称为χ^2检验法．

给定显著性水平 α，在 χ^2-分布临界值表中查自由度 $n-1$ 对

应的两个临界值 $\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)$及 $\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)$，使得

$$P\left\{\frac{nS^2}{\sigma_0^2}\geqslant\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)\right\}=\frac{\alpha}{2},$$

$$P\left\{\frac{nS^2}{\sigma_0^2}\geqslant\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)\right\}=1-\frac{\alpha}{2},$$

即

$$P\left\{\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)\leqslant\frac{nS^2}{\sigma_0^2}\leqslant\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)\right\}=1-\alpha. \quad (8.2.11)$$

当子样的观察值由(8.2.10)式算得$\frac{nS^2}{\sigma_0^2}$之值大于 $\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)$或小于$\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)$时，则否定$H_0$.

例 8.2.6 某纺织厂生产的维尼伦纤度用ξ表示，在稳定生产情况下，可假定ξ服从正态 $N(a,\sigma)$，其标准差σ按往常资料暂定为0.048. 现在随机抽取5根纤维，测得其纤度为：1.32，1.55，1.36，1.40，1.44. 试问总体ξ的方差σ^2有没有显著性的变化.

解 将上述问题归结为下述检验问题:

$$H_0:\ \sigma=0.048,\quad H_1:\ \sigma\neq0.048.$$

利用(8.2.10)式建立的统计量$\frac{nS^2}{\sigma_0^2}\sim\chi^2_{(n-1)}$对$H_0$作显著性检验.

实际计算结果如下:

$$\bar{x}=1.414,$$

$$\chi_4^2=\frac{1}{(0.048)^2}[(0.094)^2+(0.136)^2+(0.054)^2+(0.014)^2+(0.026)^2]=13.6.$$

给定显著性水平$\alpha=0.1$，在χ^2-分布临界值表中查得数值$\chi_4^2(0.05)=9.49$及$\chi_4^2(0.95)=0.711$，使得

$$P\{0.711\leqslant\chi_4^2\leqslant9.49\}=0.90.$$

由于13.6不在区间[0.711, 9.49]内，因而否定H_0，即认为方差显著地改变了.

第二种情形：两个总体的方差比较问题，此时

$$H_0:\ \sigma_1^2=\sigma_2^2,\quad H_1:\ \sigma_1^2\neq\sigma_2^2.$$

假定总体ξ及η分别服从正态$N(a_1,\ \sigma_1)$及正态$N(a_2,\sigma_2)$，现在独立地分别抽取子样$\xi_1,\cdots,\xi_{n_1}$及$\eta_1,\cdots,\eta_{n_2}$，记

$$\bar{\xi}=\frac{1}{n_1}\sum_{i=1}^{n_1}\xi_i,\qquad S_1^2=\frac{1}{n_1}\sum_{i=1}^{n_1}(\xi_i-\bar{\xi})^2;$$

$$\bar{\eta}=\frac{1}{n_2}\sum_{j=1}^{n_2}\eta_j,\qquad S_2^2=\frac{1}{n_2}\sum_{j=1}^{n_2}(\eta_j-\bar{\eta})^2.$$

则由(6.3.5)式知

$$\frac{n_1S_1^2}{\sigma_1^2}\sim\chi^2_{(n_1-1)},$$

$$\frac{n_2S_2^2}{\sigma_2^2}\sim\chi^2_{(n_2-1)}.$$

由定理6.3.2知

$$\left[\frac{\dfrac{n_1S_1^2}{\sigma_1^2}}{(n_1-1)}\bigg/\frac{\dfrac{n_2S_2^2}{\sigma_2^2}}{(n_2-1)}\right]\sim F_{(n_1-1,n_2-1)},$$

因此，当H_0成立时，则有

$$\frac{(n_2-1)n_1S_1^2}{(n_1-1)n_2S_2^2}\sim F_{(n_1-1,n_2-1)},\tag{8.2.12}$$

即$\dfrac{(n_2-1)n_1S_1^2}{(n_1-1)n_2S_2^2}$服从第一自由度$n_1-1$、第二自由度$n_2-1$的$F$-分布，且不带有未知参数，因而它可作为判断$H_0$的检验统计量. 用它作为检验统计量进行显著性检验的方法称为F检验法.

给定显著性水平α，怎样选择临界值呢？我们知道

$$E\left(\frac{n_1S_1^2}{n_1-1}\right)=\sigma_1^2,\quad E\left(\frac{n_2S_2^2}{n_2-1}\right)=\sigma_2^2.$$

若 H_0 正确，则 $F_{(n_1-1,n_2-1)}$的比值大约为 1，因此这个 F 值太大或太小，都表明 H_0 可能不正确. 一般书本上附的F分布表的临界值都是下述这种表达式所确定的:

$$P\{F \geqslant F(\alpha)\} = \alpha.$$

于是选取临界值使得

$$P\left\{F \geqslant F_{(n_1-1,n_2-1)}\left(\frac{\alpha}{2}\right)\right\} = \frac{\alpha}{2}$$

$$P\left\{F \geqslant F_{(n_1-1,n_2-1)}\left(1-\frac{\alpha}{2}\right)\right\} = 1-\frac{\alpha}{1}$$

当(8.2.12)式的F统计量的观察值大于 $F_{(n_1-1,n_2-1)}\left(\frac{\alpha}{2}\right)$或者小于 $F_{(n_1-1,n_2-1)}\left(1-\frac{\alpha}{2}\right)$时，则否定 H_0.

例 8.2.7 设用两种不同方法冶炼某种金属材料，分别抽样测定其杂质的含量(单位：百分率)数据如表 8.2.6 及表 8.2.7. 试问这两种冶炼法的杂质含量的差异性是否显著不同.

表 8.2.6

原冶炼方法：杂质含量			
26.9	22.3	27.2	25.1
22.8	24.2	30.2	
25.7	26.1	24.5	
23.0	26.4	29.5	

表 8.2.7

新冶炼方法：杂质含量		
22.6	24.3	23.4
22.5	21.9	
20.6	20.6	
23.5	23.2	

解 原方法与新方法的杂质含量分别记作 ξ 及 η，假定 ξ 服从正态 $N(a_1, \sigma_1)$，η 服从正态 $N(a_2, \sigma_2)$，于是提出如下假设检验问题：

$$H_0:\ \sigma_1=\sigma_2, H_1:\ \sigma_1\neq\sigma_2.$$

实际计算结果如下：

$$n_1=13, \bar{x}=25.68,\ S_1^2=\frac{1}{13}\sum_{i=1}^{13}(x_i-\bar{x})^2=5.411;$$

$$n_2=9, \bar{y}=22.51,\ S_2^2=\frac{1}{9}\sum_{i=1}^{9}(y_i-\bar{y})^2=1.459.$$

由(8.2.12)式知

$$F_{(12,8)}=\frac{n_1S_1^2/(n_1-1)}{n_2S_2^2/(n_2-1)}=\frac{6.173}{1.652}\approx 3.74.$$

给定显著性水平 $\alpha=0.10$，在 F 分布表中查第一自由度 12、第二自由度 8 的临界值为 $F_{(12,8)}(0.05)=3.28$，由于 $3.74>3.28$，因而否定 H_0，即认为这两种冶炼法的杂质含量的差异性是显著不同的.

§8.3 非参数的检验

现在，我们来讨论非参数的假设检验，这有两方面的问题，一是总体 ξ 的分布函数 $F(x)$ 的拟合检验，二是随机变数之间的独立性与相关性的检验.

一、分布函数的拟合检验

考虑如下假设检验问题：

$$H_0:\ F(x)=F_0(x), H_1:\ F(x)\neq F_0(x).$$

对 H_0 作显著性检验，通常称之为分布函数的拟合检验. 在此，$F_0(x)$ 为需要检验的某个已知的分布函数，$F_0(x;\theta)$ 中也可以含有未知参数 θ. 如例 8.1.2，对于呼唤次数 η，假定 η 服从泊松分布

$P(x;\lambda)$的话，那么$F_0(x)$就是泊松分布函数，λ为参数. 关于总体的分布函数，怎样提得较为准确的假设，即$F_0(x)$的函数表达式怎样提得较为准确，这可以由子样$\xi_1,\cdots,\xi_n$作经验分布函数$F_n^*(x)$，从中得到启发；也可以从第二章介绍的几种常用的概率分布的物理模型中得到启发.

对H_0作显著性检验，按不同的具体问题，可建立不同的检验统计量. 在这里，我们将介绍皮尔逊定理，它是利用子样建立的一种统计量，这个统计量的渐近分布为χ^2-分布；也介绍柯尔莫哥洛夫——斯米尔诺夫定理，它是从经验分布函数$F_n^*(x)$出发建立的统计量，这些统计量的极限分布是确定的并可计算的. 上述定理都是讨论某一统计量的极限分布问题，因而用这种统计量作显著性检验的检验统计量时，需要子样的容量n比较大，即所谓适用于大样问题. 但对于特殊的情形，如例 8.3.1 中的光通量ξ，假设ξ服从正态$N(a,\sigma)$，对于这个"假设"，可用简便的正态概率纸来作检验.

*1. 正态概率纸的应用

例 8.3.1 某工厂生产一种220伏25瓦的白炽灯泡，其光通量(单位：流明)用ξ表示，ξ为一随机变数，假设ξ是服从正态$N(a,\sigma)$，试问这个假设是否正确.

现在从总体ξ中抽取容量为$n=120$的子样(对于有限总体，即个体是有限的情形，一定要用有返回抽取方式，随机地独立抽取子样)，进行观察得光通量ξ的120个观察值，亦即随机地抽取120个灯泡测得其光通量的数据，列如下表：

表 8.3.1 白炽灯泡测试数据(单位：流明)

216	203	197	208	206	209	206	208	202	203
206	213	218	207	208	202	194	203	213	211
193	213	208	208	204	206	204	206	208	209
213	203	206	207	196	201	208	207	213	208
210	208	211	211	214	220	211	203	216	224

续表

211	209	218	214	219	211	208	221	211	218
218	190	219	211	208	199	214	207	207	214
206	217	214	201	212	213	211	212	216	206
210	216	204	221	208	209	214	214	199	204
211	201	216	211	209	208	209	202	211	207
202	205	206	216	206	213	206	207	200	198
200	202	203	208	216	206	222	213	209	219

考察如下假设检验问题:

$$H_0: F(x) = F_0(x), \quad H_1: F(x) \neq F_0(x).$$

其中 $F_0(x)$ 为正态 $N(a, \sigma)$ 的分布函数.

现在,我们用正态概率纸来对 H_0 作检验.

(1) 作直方图

把子样 $\xi_1, \cdots, \xi_n$ 的观察值 $x_1, \cdots, x_n$ 分成 m 个小组,分组的办法是:将包含 $x_1, \cdots, x_n$ 在内的某个适当的区间 (y_0, y_m) 分为互不相交的 m 个子区间 $\Delta_i = [y_{i-1}, y_i)$, $i = 1, \cdots, m$, 使得

$$y_0 < y_1 < \cdots < y_{m-1} < y_m,$$

以 ν_i 记作子样的观察值落入第 i 个小区间的频数, $i = 1, \cdots, m$. 显然, $\sum_{i=1}^{m} \nu_i = n$, 称 $\frac{\nu_i}{n}$ 为子样观察值落入第 i 个小区间的频率,以子样观察值作横坐标,以相应的频数为纵坐标,作出直方图如图 8.3.1.

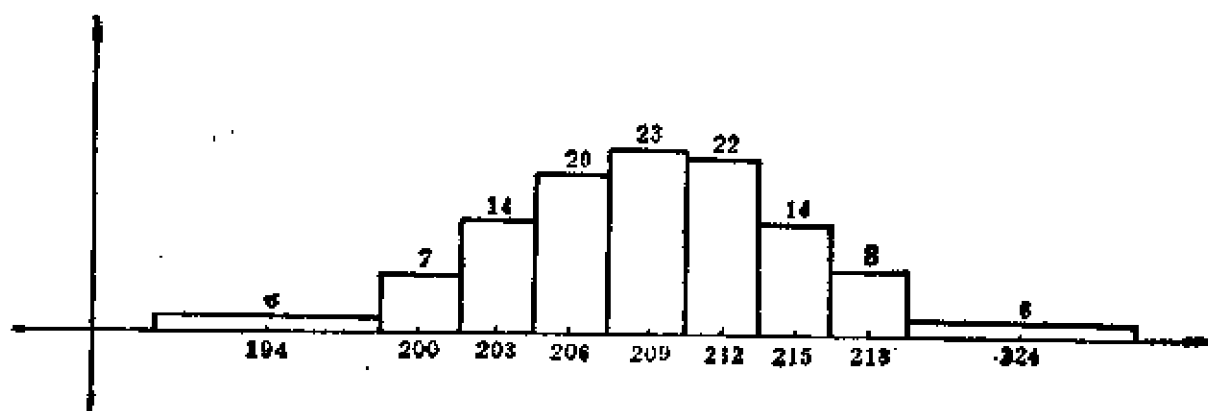


图 8.3.1

对于表 8.3.1 中的 120 个数据，可取 189.5 为下界，228.5 为上界，将 (189.5, 228.5) 按等间距离为 3 划分为 13 个小区间，发现前三个小区间及后三个小区间的 v_i 值(即频数值)都太小，应适当并小区间，使得每个小区间的 v_i 值都不小于 5. 经适当并小区间后，成为 9 个小区间，第一个小区间为 (189.5, 198.5)，第 9 个小区间为[219.5, 228.5). 第 2 至 8 这七个小区间的间距为 3. 这里假定测量精确度为个位的 1，小区间端点都带 0.5 是为了计算频数方便. 详见下表:

表 8.3.2

编号	小区间	频数	累积频数	累积频率
1	$(-\infty, 198.5)$	6	6	5%
2	[198.5, 201.5)	7	13	10.83%
3	[201.5, 204.5)	14	27	22.5%
4	[204.5, 207.5)	20	47	39.17%
5	[207.5, 210.5)	23	70	58.33%
6	[210.5, 213.5)	22	92	76.66%
7	[213.5, 216.5)	14	106	88.33%
8	[216.5, 219.5)	8	114	95%
9	$[219.5, \infty)$	6	120	100%

(2) 正态概率纸

由子样观察值所作的直方图如图 8.3.1，可作为要考察的连续型随机变数的密度曲线 $f(x)$ 的粗略反映，从图 8.3.1 看出，它呈现了中间高、两头低、左右基本对称的特性，很象正态分布的密度函数曲线. 试问"总体 ξ 服从正态分布"这个结论成立吗?现在，我们利用一种特殊的坐标纸——正态概率纸来进行检验，这是一种简便的判断准则.

正态概率纸是这样的：设 ξ 服从正态 $N(0,1)$，其分布函数为

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du,$$

用随机变数 ξ 的观察值 x 为横坐标，$\Phi(x)$ 为纵坐标，要使 $(x_1, \Phi(x_1))$，$(x_2, \Phi(x_2))$，…，$(x_n, \Phi(x_n))$ 这 n 个点在一条直线上，应该这样建立坐标系：横坐标轴按等间隔作刻度，依次记作 $x_1, \cdots, x_n$，而纵坐标轴上的刻度不等间隔，依

次记作 $\Phi(x_1),\cdots,\Phi(x_n)$，用这样特殊的坐标系为坐标纸，称之为正态概率纸. 由此可见，若总体 ξ 是服从正态分布，其分布函数为

$$F(x;a,\sigma)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(u-a)^2}{2\sigma^2}}du,$$

那么，以 $(x,F(x))$ 为点，在正态概率纸上作图时，便得一条直线. 反之，在正态概率纸上的任一条直线，对应着一个正态分布的分布函数. 总之，在正态概率纸上，一条直线同正态分布函数 $F(x;a,\sigma)$ 是一一对应的. 例如 σ 取不同值，对应着不同的正态分布函数，在正态概率纸上对应于斜率不同的直线.

(3) 估计与检验

对于例 8.3.1，按表 8.3.2 中的数据，以各小区间的中点值为横坐标，以相对应的累积频率为纵坐标，在正态概率纸上标出一个点，共能算出 9 个点，如图 8.3.2 所示. 这九个点没有集中在一条直线上. 如果总体 ξ 是服从正态分布，而这个子样确实来自这个总体，并能在很大程度上反映这个总体的概率统计特性，那么这些点应该集中在一条直线上. 诚然，子样本身是随机向量，因而总会有一些偏差，这是允许的，但偏差不能过大. 如果偏差过大或很不象一条直线，那么就应怀疑总体 ξ 是否服从正态分布. 兼顾上述的九个点，划一条直线，记为 L. 一般的说，中间的点离直线的偏差不能过大，两端的点偏差可以允许大一些，这是因为服从正态分布的随机变数 ξ 以概率 68.3% 取值于数学期望 $E(\xi)=a$ 为中心、区间长度为 σ 的区间内. 从图 8.3.2 看到，所有各点离直线偏差不大，因而可认为光通量 ξ 是服从正态分布.

从图 8.3.2 上还可粗略地给出数学期望 $E(\xi)=a$ 及方差 $D(\xi)=\sigma^2$ 的估计值. 我们知道，正态分布的密度函数，是关于数学期望对称的，所以在直线 L 上是以纵坐标为 0.5 的点所对应的横坐标，即为 $E(\xi)=a$. 从图 8.3.2 看到，a 的估计值约为 209. 又因为事件 $\{\xi<a-\sigma\}$ 发生的概率为 0.1587，所以在直线 L 上以 0.1587 为纵坐标的点对应的横坐标为 $a-\sigma$. 从图 8.3.2 看到，$a-\sigma$ 这个值约为 202.5，于是 σ 的估计值为：

$$209-202.5=6.5.$$

附：正态概率纸

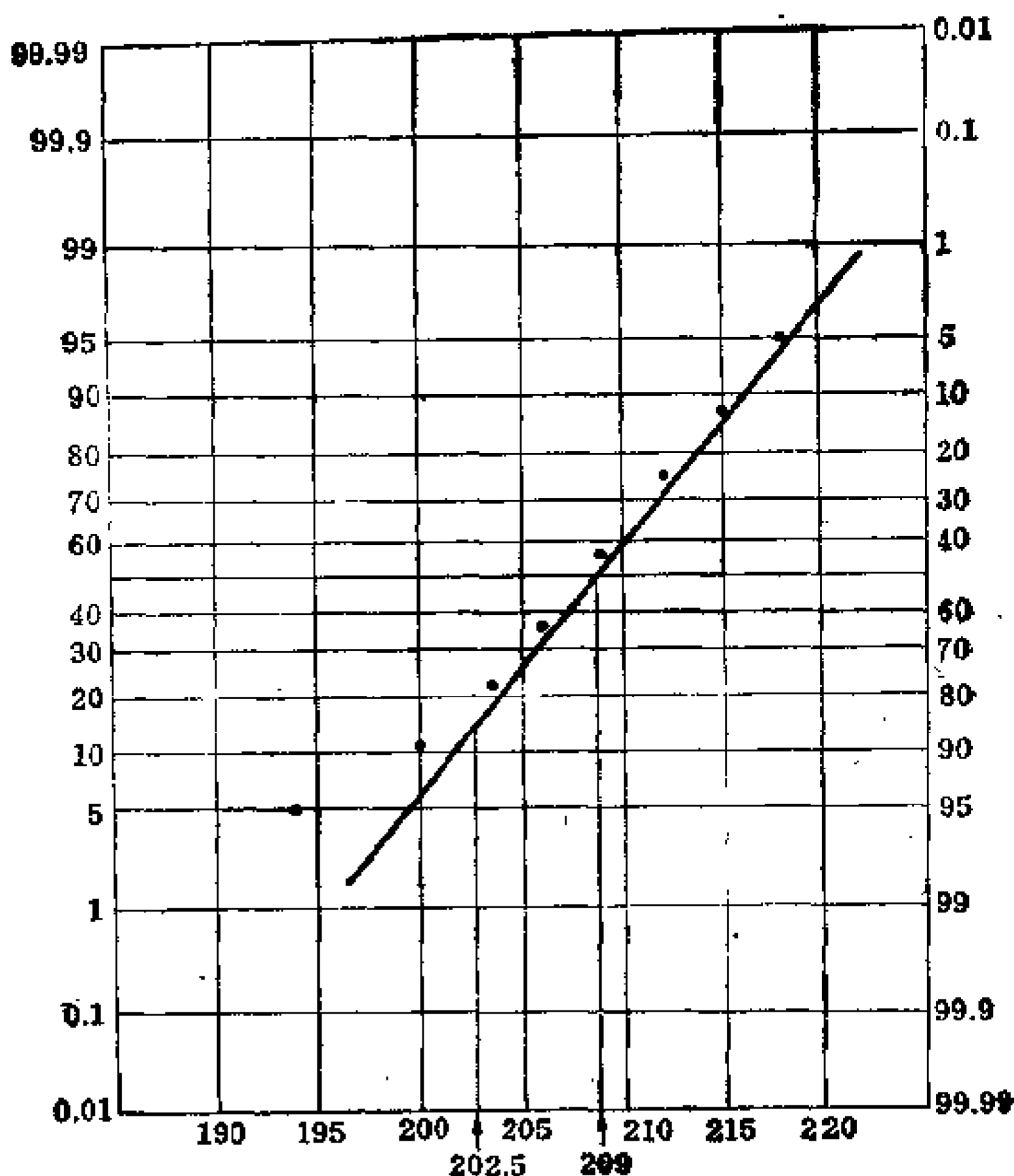


图 8.3.2

2. 皮尔逊 χ^2 检验

正态概率纸的应用，只限于判断检验对象 $F_0(x)$ 为正态分布这个假设是否正确. 对于较为一般的检验对象 $F_0(x)$，如 $F_0(x)$为泊松分布、均匀分布等情形，需要建立较为通用的检验统计量. 为此，下面介绍皮尔逊定理，这个定理所建立的统计量的极限分布为 χ^2-分布，不论总体 ξ 的分布函数 $F(x)$是什么类型，定理的结论都

适用于作检验判断，这种检验法称之为皮尔逊 χ^2 检验.

我们接着上面的例 8.3.1 作讨论. 从表 8.3.2 看到，v_i 为子样 $\xi_1, \cdots, \xi_n$ 中落入第 i 个小区间的频数(即个数)，称 v_i 为观察频数. 如果 H_0 成立，由给定的分布函数 $F_0(x)$，计算得到：

$$p_i = F_0(y_i) - F_0(y_{i-1}),$$

其中 $0<p_i<1$, $\sum_{i=1}^{m} p_i = 1, i=1, \cdots, m$，称 np_i 为子样 $\xi_1, \cdots, \xi_n$ 落入第 i 个小区间的理论频数. 由此可见，v_i 依赖于子样 $\xi_1, \cdots, \xi_n$ 的观察值. 考察统计量

$$\eta = \sum_{i=1}^{m} \frac{(v_i - np_i)^2}{np_i} = \sum_{i=1}^{m} \frac{v_i^2}{np_i} - n, \tag{8.3.1}$$

其中 η 依赖于 v_i 及 p_i，因而它与 $F_0(x)$ 建立了一定的关系，它可作为判断 H_0 的检验统计量.

我们来分析(8.3.1)式所建立的统计量 η 的实际意义. 当 H_0 成立时，由第五章的波雷尔强大数定律知道，以概率为 1 地使得 $\frac{v_i}{n} \to p_i$，即当 n 足够大时 $v_i \approx np_i$，则 η 以概率为 1 地取值于零的附近，因而由(8.3.1)式所建立的统计量 η 可以用来判断 $F(x)$ 与 $F_0(x)$ 之间的差异性是否显著. 正因为这样，这个统计量 η（它不带有未知参数)可作为判断 H_0 的检验统计量. 试问，统计量 η 服从什么分布？由于统计量 η 依赖于 m 及 n，如果固定 m，当 $n \to \infty$ 时，η 的极限分布是什么？皮尔逊于 1900 年证明了如下定理.

皮尔逊定理 不论 $F_0(x)$ 是什么分布，当 H_0 正确时，则由(8.3.1)式建立的统计量 η 以自由度 $m-1$ 的 χ^2-分布为极限分布，其中 $F_0(x)$ 不带有未知参数.

证 我们就 $m=2$ 的情形先作证明.

$$\begin{aligned}\eta &= \frac{(v_1-np_1)^2}{np_1}+\frac{(v_2-np_2)^2}{np_2}\\ &= \frac{(v_1-np_1)^2}{np_1}+\frac{[(n-v_1)-n(1-p_1)]^2}{n(1-p_1)}\\ &= \frac{(v_1-np_1)^2}{np_1}+\frac{(v_1-np_1)^2}{n(1-p_1)}\\ &= \frac{(v_1-np_1)^2}{np_1(1-p_1)}. \end{aligned} \tag{8.3.2}$$

引进记号

$$v_1=\sum_{k=1}^{n}\zeta_k,$$

其中 $\zeta_1,\cdots,\zeta_n$ 是 n 个相互独立的如下定义的随机变数:

$$\zeta_k=\begin{cases}1, & \xi_k\in[y_0,y_1];\\ 0, & \xi_k\overline{\in}[y_0,y_1],\end{cases}\quad k=1,\cdots,n,$$

则

$$P\{\zeta_k=1\}=F_0(y_1)-F_0(y_0)=p_1;$$

$$E(v_1)=np_1,\qquad D(v_1)=np_1(1-p_1).$$

由定理 5.4.3 知,统计量

$$\frac{v_1-np_1}{\sqrt{np_1(1-p_1)}}$$

的极限分布为 $N(0,1)$,因而统计量 η 的极限分布为 χ_1^2.

对于一般 m 的情形,定理的详细证明用小字体叙述如下:

记

$$u_i=\frac{v_i-np_i}{\sqrt{np_i}}\qquad i=1,\cdots,m. \tag{8.3.3}$$

(1) 设 H_0 正确, 即总体 ξ 的分布函数为 $F_0(x)$, 则 $(v_1,\cdots,v_m)$ 有 m 项分布:

$$P\{v_1=k_1,\cdots,v_m=k_m\}=\frac{n!}{k_1!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m},$$

其中 v_i 为随机变数，k_i 为 v_i 的观察值（$i=1,\cdots,m$），每一个 k_i 都是非负整数，且 $\sum_{i=1}^{m}k_i=n$. $(v_1,\cdots,v_m)$ 的特征函数为

$$\varphi_v(t_1,\cdots,t_m)=(p_1e^{jt_1}+\cdots+p_me^{jt_m})^n,$$

那么(8.3.3)式中 $(u_1,\cdots,u_m)$ 的特征函数为：

$$\varphi_u(t_1,\cdots,t_m)=e^{-j\sqrt{n}\sum_{i=1}^{m}t_i\sqrt{p_i}}(p_1e^{jt_1/\sqrt{np_1}}+\cdots+p_me^{jt_m/\sqrt{np_m}})^n.$$

利用台劳展开，则有

$$\ln\varphi_u(t_1,\cdots,t_m)=-j\sqrt{n}\sum_{i=1}^{m}t_i\sqrt{p_i}+n\ln\left[\sum_{i=1}^{m}p_i(e^{jt_i/\sqrt{np_i}}-1)+1\right]$$

$$=-j\sqrt{n}\sum_{i=1}^{m}t_i\sqrt{p_i}+n\left[\sum_{i=1}^{m}p_i(e^{jt_i/\sqrt{np_i}}-1)\right]$$

$$-\frac{n}{2}\left[\sum_{i=1}^{m}p_i(e^{jt_i/\sqrt{np_i}}-1)\right]^2+o(1)$$

$$=-j\sqrt{n}\sum_{i=1}^{m}t_i\sqrt{p_i}+n\sum_{i=1}^{m}p_i\left(\frac{jt_i}{\sqrt{np_i}}-\frac{t_i^2}{2np_i}\right)$$

$$-\frac{n}{2}\left[\sum_{i=1}^{m}p_i\left(\frac{jt_i}{\sqrt{np_i}}-\frac{t_i^2}{2np_i}\right)\right]^2+o(1)$$

$$=-\frac{1}{2}\sum_{i=1}^{m}t_i^2-\frac{n}{2}\left(\sum_{i=1}^{m}jt_i\sqrt{\frac{p_i}{n}}\right)^2+o(1)$$

$$=-\frac{1}{2}\sum_{i=1}^{m}t_i^2+\frac{1}{2}\left(\sum_{i=1}^{m}t_i\sqrt{p_i}\right)^2+o(1),$$

于是

$$\lim_{n\to\infty}\varphi_u(t_1,\cdots,t_m)=\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m}t_i^2-\left(\sum_{i=1}^{m}t_i\sqrt{p_i}\right)^2\right]\right\}.$$

(2) 作正交变换如下

$$s_l=\sum_{i=1}^{m}c_{li}t_i\qquad(l=1,\cdots,m-1),$$

$$s_m=\sum_{i=1}^{m} t_i\sqrt{p_i}.$$

由正交变换的性质知

$$\sum_{i=1}^{m} t_i^2=\sum_{i=1}^{m} s_i^2,$$

于是

$$\lim_{n\to\infty}\varphi_n(t_1,\cdots,t_m)=\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{m} s_i^2-s_m^2\right]\right\}=e^{-\frac{1}{2}\sum\limits_{j=1}^{m-1} s_j^2}.$$

右端是 $m-1$ 维正态分布的特征函数，而且有此分布的 $m-1$ 维随机变数是不相关的，即右端是 $m-1$ 个相互独立的且都服从正态分布的随机变数的特征函数之积. 由特征函数的逆极限定理知，当 $n\to\infty$ 时，$(u_1,\cdots,u_m)$ 的极限分布是 $m-1$ 维正态分布（逆极限定理对多维分布也成立）. 从而推知，当 $n\to\infty$ 时，统计量 $\eta=\sum\limits_{i=1}^{m} u_i^2$ 的极限分布是 $\chi^2_{(m-1)}$-分布. ∎

给定了显著性水平 α，对于(8.3.1)式建立的检验统计量 η，怎样选择临界值确定 H_0 的否定域？由于 η 的精确分布不知道，我们是用它的极限分布 $\chi^2_{(m-1)}$ 近似地选择临界值. 同时还应注意到，(8.3.1)式的 χ^2 检验同(8.2.10)式及(8.2.11)式选择临界值及否定域是有所不同的. 粗略地看到，若(8.3.1) 式计算统计量 η 的观察值很大，H_0 很可能被否定. 因此，应该这样选择临界值 $\chi^2_{(m-1)}(\alpha)$，使得

$$P\{\eta>\chi^2_{(m-1)}(\alpha)\}=\alpha. \tag{8.3.4}$$

即否定域为区间 $(\chi^2_{(m-1)}(\alpha),\infty)$. 就是说，由(8.3.1)式计算得统计量 η 的观察值，若大于临界值 $\chi^2_{(m-1)}(\alpha)$，则在显著性水平 α 下否定 H_0.

如果 H_0 中的 $F_0(x;\theta_1,\cdots,\theta_l)$ 含有 l 个未知参数，则首先用这 l 个未知参数的极大似然法估计量 $\hat{\theta}_1,\cdots,\hat{\theta}_l$ 来代替 $\theta_1,\cdots,\theta_l$ 使 $F_0(x;\hat{\theta}_1,\cdots,\hat{\theta}_l)$ 不含未知参数，然后应用(8.3.1)式建立的统计量

$\eta(\hat{\theta}_1, \cdots, \hat{\theta}_l)$，再用上面介绍的方法进行显著性检验．但注意此时以 $\chi^2_{(m-1-l)}$-分布为其极限分布，其中 $m>1+l$．(参看[7]中第30章的 30.3．)

例 8.3.2 对例 8.3.1 应用皮尔逊 χ^2 检验．此时

$$H_0:\ F(x)=F_0(x),\ H_1:\ F(x)\neq F_0(x).$$

其中 $F_0(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(u-a)^2}{2\sigma^2}}du$，$a$ 及 σ 都是未知参数．

解 利用表 8.3.2 中的数据，由(8.3.1)式计算统计量 η 的观察值，判断 H_0 是否成立．关键在于计算 np_i，其中

$$p_i=F_0(y_i)-F_0(y_{i-1})\quad (i=1,\cdots,m).$$

我们知道，对于正态分布，a 及 σ^2 的极大似然法估计量为：

$$\hat{a}=\bar{\xi},\ \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2.$$

现在用表 8.3.1 中的数据 x_i 求出 $\bar{\xi}$ 及 S_n^2 的观察值 $\bar{x}$ 及 s^2，作为 a 及 σ^2 的估计值．

表 8.3.3

编　号	$[y_{i-1}, y_i)$	v_i	np_i
1	$(-\infty, 198.5)$	6	6.1
2	$[198.5, 201.5)$	7	8.7
3	$[201.5, 204.5)$	14	14.5
4	$[204.5, 207.5)$	20	19.7
5	$[207.5, 210.5)$	23	21.8
6	$[210.5, 213.5)$	22	19.7
7	$[213.5, 216.5)$	14	14.5
8	$[216.5, 219.5)$	8	8.8
9	$[219.5, \infty)$	6	6.1

计算得

$$\hat{a}=\bar{x}\approx 209,$$

$$\hat{\sigma}^2 = s^2 \approx 42.77,$$

$$\hat{\sigma} \approx 6.5.$$

因此 $F_0(x)$为正态 $N(209, 6.5)$的分布函数，可算得

$$\begin{aligned}
p_1 &= F_0(198.5) - F_0(-\infty) \\
&= P\{-\infty < \xi < 198.5\} \\
&= P\left\{-\infty < \frac{\xi - 209}{6.5} < -1.62\right\} \\
&= \Phi(-1.62) - \Phi(-\infty) = 1 - \Phi(1.62) \\
&= 1 - 0.94738 = 0.05262, \\
p_2 &= F_0(210.5) - F_0(198.5) \\
&= P\{198.5 \leqslant \xi < 201.5\} \\
&= P\left\{-1.62 \leqslant \frac{\xi - 209}{6.5} < -1.15\right\} \\
&= \Phi(-1.15) - \Phi(-1.62) \\
&= \Phi(1.62) - \Phi(1.15) \\
&= 0.94738 - 0.87493 = 0.07245,
\end{aligned}$$

其中 $\Phi(x)$为标准正态 $N(0, 1)$的分布函数，类似于 p_1 及 p_2 的算法，可逐一计算 $p_3, \cdots, p_9$ 各值如表 8.3.4 所示.

表 8.3.4

编　号	v_i	v_i^2	v_i^2/np_i
1	6	36	5.701
2	7	49	5.568
3	14	196	13.517
4	20	400	20.305
5	23	529	24.266
6	22	484	24.568
7	14	196	13.448
8	8	64	7.273
9	6	36	5.701

因而统计量η的观察值为

$$\eta=\sum_{i=1}^{9}(v_i^2/np_i)-120\approx 0.347.$$

给定显著性水平$\alpha=0.05$，由于自由度等于$9-1-2=6$，查得临界值为$\chi_6^2(0.05)=12.59$，由于$0.347<12.59$，所以不否定H_0，在实际工作中可认为光通量ξ服从正态$N(209,6.5)$.

例 8.3.3 某箱子中盛有 10 种球，现在从中有返回地随机抽取 200 个，其中第i种球共取得v_i个$(i=1,\cdots,10)$，数据记录在表 8.3.5 中.

表 8.3.5

种 别	v_i	v_i^2	np_i	v_i^2/np_i
1	35	1225	20	61.25
2	16	256	20	12.8
3	15	225	20	11.25
4	17	289	20	14.45
5	17	289	20	14.45
6	19	361	20	18.05
7	11	121	20	6.05
8	16	256	20	12.8
9	30	900	20	45
10	24	576	20	28.8
$\sum_{i=1}^{10}$	200		200	224.9

令 H_0: 箱子中各种球的个数相同，H_1: 至少有两种球的个数不同. 检验这个假设是否正确.

解 此时$m=10$, $y_0=-\infty$, $y_i=i$ $(i=1,\cdots,10)$, $y_{11}=\infty$. 若H_0正确，采用有返回抽取方式抽得每种球的概率都相同，皆为$\frac{1}{10}$,

亦即用有返回抽取方式抽取一球为第 i 种球的概率 $p_i=\frac{1}{10}(i=1,\cdots,10)$. 又 $n=200$, 则 $np_i=20$ $(i=1,\cdots,10)$, 并记录在表 8.3.5 中.

给定显著性水平 $\alpha=0.05$, 由自由度为 $10-1=9$ 的χ^2-分布表查得临界值 $\chi_9^2(0.05)=23.6$. 由于 η 的观察值为 $224.9-200=24.9>23.6$, 所以否定 H_0. 事实上, 从表 8.3.5 中 v_i 的观察数据看到, 第 1 及第 9 这两种的个数明显地较多.

例 8.3.4 卢瑟福(Rutherford, 1920 年)与盖革(Geiger)作了一个著名的实验①, 他们观察了长为7.5秒的时间间隔里到达某个计数器的由某块放射物质放出的 α 质点数,, 共观察了 $n=2608$ 次. 表 8.3.6 中的第 1 列给出的是质点数 i, 第 2 列表示有 i 个质点到达计数器的时间间隔数 v_i(每个时间间隔都是长为 7.5秒). 试问这种分布规律是否服从泊松分布? 即考察如下检验问题:

$$H_0:\ F(x)=F_0(x),\ H_1:\ F(x)\neq F_0(x).$$

其中$F_0(x)$是泊松分布

$$P(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}\quad x=0,1,2,\cdots$$

的分布函数, $\lambda>0$ 为其参数.

解 从第二章的 §2.1 中讲到的满足泊松分布的四个条件来看, 可以认为原假设 H_0 是正确的. 现利用 χ^2 检验法对 H_0作出判断, 主要是看一看理论频数 np_i 与观察频数 v_i 之间的差异性是大还是小. 参数 λ 的极大似然法估计值为:

$$\lambda=\frac{1}{n}\sum_i iv_i$$

① Rutherford Chadwick and Ellis, Radiations from Radivative Substances Cambridge 1920. p. 172.

$$=\frac{10086}{2608}=3.87.$$

表 8.3.6 中第 3 列的数值np_i，利用泊松分布表计算p_i值而得到，其中

$$p_i=\frac{(3.87)^i}{i!}e^{-3.87}\quad i=0,1,2,\cdots.$$

如 $p_0=0.020824, np_0=54.309.$

表 8.3.6

i	v_i	np_i
0	57	54.309
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
⩾10	16	17.075
	2608	2608.000

读者可以象表 8.3.4 那样计算 v_i^2 及 v_i^2/np_i 各值。应当注意到，在用 χ^2 检验法时，自由度为 $m-1-1$，而 $m=11$，在给定显著性水平 α 的时候，查自由度 9 所对应的临界值 $\chi_9^2(\alpha)$，H_0 的否定域为 $(\chi_9^2(\alpha),\infty)$。

例 8.3.5 在表 8.3.7 中给出了在 $n=267$ 次观察中电话接错次数 i 的观察频数 v_i 及理论频数 np_i 各值①。提出同例 8.3.4

① F. Thorndike, Applications of Poisson's Probability Summation, *The Bell System Technical Journal*, Vol. 5(1926).

那样的检验问题.

表 8.3.7

i	v_i	np_i
0—2	1	2.05
3	5	4.76
4	11	10.39
5	14	18.16
6	22	26.45
7	43	33.03
8	31	36.09
9	40	35.04
10	35	30.63
11	20	24.34
12	18	17.72
13	12	11.92
14	7	7.44
15	6	4.33
≥16	2	4.65
	267	267.00

解 参数 λ 的极大似然法估计值为

$$\hat{\lambda}=\frac{1}{n}\sum_i i v_i=8.74.$$

在用 χ^2 检验法时，自由度为 $15-1-1=13$. 给定显著性水平 $\alpha=0.05$ 时，查得 $\chi^2_{13}(0.05)=22.362$.

***二、不相关与独立性的检验**

我们讨论如下两种情形的非参数性的检验问题.

1. 不相关的检验

设 (ξ,η) 服从如例2.2.4所叙述的二维正态分布，即 ξ 服从正态 $N(a_1,\sigma_1)$，η 服从正态 $N(a_2,\sigma_2)$，ρ 为 ξ 与 η 之间的相关系数. 提出如下假设检验问题:

$$H_0:\ \rho=0,\ H_1:\ \rho\neq 0.$$

在例 7.1.4 中给出了相关系数 ρ 的矩法估计量 R:

$$R=\frac{S_{12}}{S_1\cdot S_2}.$$

统计量 R 是两个子样之间的相关系数,它不带有未知参数 ρ,可作为 H_0 的检验统计量. 费歇于 1915 年导出了统计量 R 的密度函数(参看[7]的 29.7)为:

$$f_R(r)=\frac{n-2}{\pi}(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}$$

$$\int_0^1\frac{x^{n-2}}{(1-\rho rx)^{n-1}}\frac{dx}{\sqrt{1-x^2}},$$

$$|r|<1,$$

其中 r 为统计量 R 的观察值. 由此看到, R 的分布只与 ρ 及子样容量 n 有关.

当 H_0 成立时, R 的密度函数可简化为:

$$f_R(r)=\frac{n-2}{\pi}(1-r^2)^{\frac{n-4}{2}}\int_0^1\frac{x^{n-2}dx}{\sqrt{1-x^2}}.$$

记 $u=x^2$, 得

$$f_R(r)=\frac{n-2}{2\pi}(1-r^2)^{\frac{n-4}{2}}\int_0^1 u^{\frac{n-1}{2}-1}(1-u)^{\frac{1}{2}-1}du$$

$$=\frac{n-2}{2\pi}(1-r^2)^{\frac{n-4}{2}}B\left(\frac{n-1}{2},\ \frac{1}{2}\right)$$

$$=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}(1-r^2)^{\frac{n-4}{2}}\quad(|r|<1). \tag{8.3.5}$$

F. N. David(1938 年)利用(8.3.5)式, 对于给定的显著性水平 α 及不同的子样容量 n,编造了临界值表, 便于对 H_0 作显著性检验之用(参看[5]的第 9 章 § 9 及[25]的三).

下面给出对 H_0 作显著性检验的另一个检验统计量. 我们考虑统计量

$$T=\sqrt{n-2}\frac{R}{\sqrt{1-R^2}}, \tag{8.3.6}$$

其中统计量 R 的密度函数由(8.3.5)给出，试求 T 的密度函数 $g(t)$ 的表达式. 根据 §2.4 求随机变数的函数的分布函数方法，得到:

$$g(t)=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}\left(1+\frac{t^2}{n-2}\right)^{-\frac{n-4}{2}}\frac{1}{\sqrt{n-2}}\left(1+\frac{t^2}{n-2}\right)^{-\frac{3}{2}}$$

$$=\frac{1}{\sqrt{(n-2)\pi}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\left(1+\frac{t^2}{n-2}\right)^{-\frac{n-1}{2}} \tag{8.3.7}$$

(令 $t=\sqrt{n-2}\frac{r}{\sqrt{1-r^2}}$, 则 $1-r^2=\left(1+\frac{t^2}{n-2}\right)^{-1}$, 由(8.3.5)即得(8.3.7). 由(8.3.7)式给出的 $g(t)$ 就是 §2.4 中所定义的自由度为 $n-2$ 的 t 分布的密度函数，因此 (8.3.6) 式的统计量 T 在 H_0 成立时服从自由度为 $n-2$ 的 t 分布，于是可用 (8.3.6) 的统计量 T 对 H_0 作显著性检验，这也是 t 检验法.

2. 独立性检验

我们知道，当(ξ,η)服从二维正态分布时，ξ 与 η 相互独立同 $\rho=0$ 是等价的，因此检验 ξ 与 η 是否相互独立，等价于 $\rho=0$ 的显著性检验. 在一般情形，这个等价性就不成立了.

我们知道，对于任何两个随机变数 ξ 及 η，ξ 与 η 相互独立是指：若(ξ,η)的联合分布函数为 $F(x,y)$，ξ 及 η 的边沿分布函数为 $F_1(x)$及 $F_2(y)$，则对一切 x 及 y 下式成立:

$$F(x,y)=F_1(x)\cdot F_2(y).$$

因此，两个随机变数之间独立性的检验问题应写为:

$$H_0:\ F(x,y)=F_1(x)F_2(y),\ H_1: F(x,y)\neq F_1(x)F_2(y).$$

在实际工作中如果总体的特征是由两个指标 ξ 及 η 联合来反映，这就要考虑二维随机变数 (ξ,η) 了. 从这个总体中，随机抽取容量为 n 的二维子样 $(\xi_1,\eta_1),\cdots,(\xi_n,\eta_n)$，将 ξ 及 η 的可能取值的范围分别分成 r 个及 s 个互不相交的小区间，用 n_{ik} 表示这个子样中"ξ 属于第 i 个小区间、η 属于第 k 个小区间"的个数 $(i=1,\cdots,r;k=1,\cdots,s)$. 记:

$$n_{i\cdot}=\sum_{k=1}^{s}n_{ik},\quad n_{\cdot k}=\sum_{i=1}^{r}n_{ik},\quad n=\sum_{i=1}^{r}\sum_{k=1}^{s}n_{ik}.$$

下面的表 8.3.8 称为联列表

表 8.3.8

i \ n_{ik} \ k	1	2	$\cdots$	s	$n_{i\cdot}$
1	n_{11}	n_{12}	$\cdots$	n_{1s}	$n_{1\cdot}$
2	n_{21}	n_{22}		n_{2s}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
r	n_{r1}	n_{r2}	$\cdots$	n_{rs}	$n_{r\cdot}$
$n_{\cdot k}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot s}$	n

从总体 (ξ,η) 中任意抽取一个元素(即一个个体)，它的"第一坐标 x 属于第 i 个区间、第二坐标 y 属于第 k 个小区间"这一事件的概率记为 $p(i,k)$，以 $p(i,\cdot)$ 及 $p(\cdot,k)$ 分别记相应的边沿概率. 则有:

$$p(i,\cdot)=\sum_{k=1}^{s}p(i,k),\ p(\cdot,k)=\sum_{i=1}^{r}p(i,k),$$

$$\sum_{i=1}^{r}\sum_{k=1}^{s}p(i,k)=\sum_{i=1}^{r}p(i,\cdot)=\sum_{k=1}^{s}p(\cdot,k)=1.$$

此时，独立性的检验就等价于检验下述原假设：

$$H_0:\ p(i,k)=p(i,\cdot)p(\cdot,k),$$

$i=1,\cdots,r;k=1,\cdots,s.$

现在，要建立 H_0 的检验统计量．首先，我们按极大似然法确定未知参数 $p(i,\cdot)$ 及 $p(\cdot,k)$ 的估计量．由于

$$p(r,\cdot)=1-\sum_{i=1}^{r-1}p(i,\cdot),$$

$$p(\cdot,s)=1-\sum_{k=1}^{s-1}p(\cdot,k),$$

所以实际上要求 $r+s-2$ 个未知参数的估计量．当 H_0 成立时，未知参数的似然函数为：

$$L=\prod_{i=1}^{r}\prod_{k=1}^{s}[p(i,k)]^{n_{ik}}=\prod_{i=1}^{r}\prod_{k=1}^{s}[p(i,\cdot)]^{n_{ik}}[p(\cdot,k)]^{n_{ik}}$$

$$=\left[\prod_{i=1}^{r}p(i,\cdot)^{n_{i\cdot}}\right]\left[\prod_{k=1}^{s}p(\cdot,k)^{n_{\cdot k}}\right]$$

$$=\left[1-\sum_{i=1}^{r-1}p(i,\cdot)\right]^{n_{r\cdot}}\left[1-\sum_{k=1}^{s-1}p(\cdot,k)\right]^{n_{\cdot s}}$$

$$\times\left[\prod_{i=1}^{r-1}p(i,\cdot)\right]^{n_{i\cdot}}\left[\prod_{k=1}^{s-1}p(\cdot,k)\right]^{n_{\cdot k}}.$$

由似然方程

$$\begin{cases}\dfrac{\partial\ln L}{\partial p(i,\cdot)}=0, & i=1,\cdots,\ r-1,\\ \dfrac{\partial\ln L}{\partial p(\cdot,k)}=0, & k=1,\cdots,s-1\end{cases}$$

解得

$$\hat{p}(i,\cdot)=\frac{n_{i\cdot}}{n},\qquad \hat{p}(\cdot,k)=\frac{n_{\cdot k}}{n},$$

$i=1,\cdots,r-1$；$k=1,\cdots,s-1$，则 $\hat{p}(i,\cdot)$ 及 $\hat{p}(\cdot,k)$ 分别为 $p(i,\cdot)$ 及 $p(\cdot,k)$ 的极大似然法估计量.

利用(8.3.1)式，导出 H_0 的检验统计量

$$\eta=\sum_{i=1}^{r}\sum_{k=1}^{s}\left[n_{ik}-\frac{n_{i\cdot}n_{\cdot k}}{n}\right]^2\Bigg/\frac{n_{i\cdot}n_{\cdot k}}{n}$$

$$=n\sum_{i=1}^{r}\sum_{k=1}^{s}\left[n_{ik}-\frac{n_{i\cdot}n_{\cdot k}}{n}\right]^2\Bigg/n_{i\cdot}n_{\cdot k},\qquad(8.3.8)$$

其中 $n\hat{p}(i,k)=n\hat{p}(i,\cdot)\hat{p}(\cdot,k)=n_{i\cdot}n_{\cdot k}/n$. 自由度为 $rs-1-(r+s-2)=(r-1)(s-1)$. 可见，当 H_0 为真时，(8.3.8)所建立的统计量 η 的极限分布为 $\chi^2_{(r-1)(s-1)}$. 给定显著性水平 α，在 χ^2-分布临界值表中查自由度 $(r-1)(s-1)$ 所对应的临界值 $\chi^2_{(r-1)(s-1)}(\alpha)$，当由(8.3.8)式计算得 η 的数值大于这个临界值时，则在显著性水平 α 下否定 H_0.

特别地，当 $r=s=2$ 时，统计量为

$$\eta=\frac{n[n_{11}n_{22}-n_{12}n_{21}]^2}{n_{1\cdot}n_{2\cdot}n_{\cdot1}n_{\cdot2}}.$$

η 的极限分布是自由度为 1 的 χ^2-分布.

§8.4 最佳检验与无偏检验

假设检验的一般理论，在数理统计学上占有重要的地位. 以上我们讨论了参数及非参数两方面的显著性检验问题. 但是，对假设检验问题还未作深入的研究. 例如对于§8.2 中所讨论到的参数检验问题，可以发现，还存在一系列的问题有待解决. 例如，当原假设不真时而作出判断不否定 H_0 的可能性有多大；为什么要选择那样的检验统计量；给定了显著性水平 α，确定 H_0 的否定域时有多少种方法；怎样给定检验法的好坏准则，而最优的检验法在什

么条件下存在，又怎样寻找. 要解决这些问题，必须引入一些新的概念. 在这一节，我们仅以参数的假设检验问题为例，讨论无偏检验及最优势检验(简称作最佳检验).

一、最佳检验

我们从两个例子的讨论，将引入新的概念.

例 8.4.1 设总体 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是未知参数，现在提出如下检验问题:

$$H_0:\ a=0,\qquad H_1: a\neq 0.$$

当原假设成立时，由(8.2.2)式知，建立了统计量

$$T=\sqrt{n-1}\frac{\bar{\xi}}{S}\sim t_{(n-1)},$$

它可作为对 H_0 的显著性检验的检验统计量. 给定显著性水平 α，可选临界值 $t_{(n-1)}(\alpha)$，使得

$$P\{|T|>t_{(n-1)}(\alpha)\}=\alpha.$$

记

$$B_0=(-\infty,-t_{(n-1)}(\alpha))\cup(t_{(n-1)}(\alpha),\infty),\qquad (8.4.1)$$

则 B_0 是 H_0 的显著性水平为 α 的否定域. 我们也可这样选择临界值 δ，使得

$$P\{|T|<\delta\}=\alpha.$$

记

$$B_\delta=(-\delta,\delta),\qquad (8.4.2)$$

B_δ 也是 H_0 的显著性水平 α 的否定域.

如图 8.4.1 所示，$g(t)$ 为检验统计量 T 的密度函数.

由此可见，在参数的显著性检验中，对于相同的检验统计量及相同的显著性水平 α，可以有不同的临界值，相应地有不同的否定域.

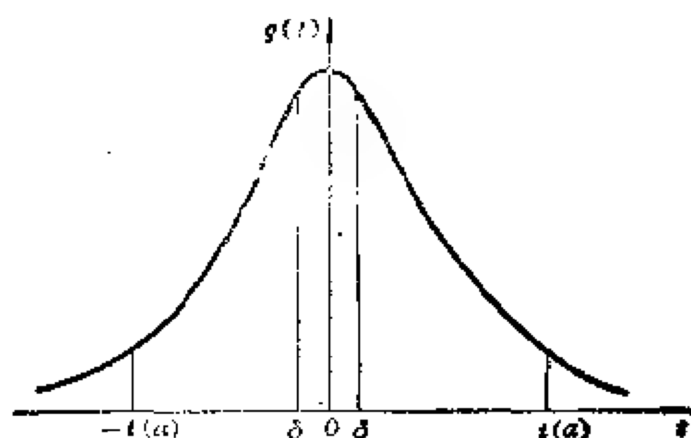


图 8.4.1

例 8.4.2 设总体 ξ 服从正态 $N(a, \sigma_0)$，其中 σ_0 为已知，考虑如下检验问题：

$$H_0:\ a=a_0,\quad H_1:\ a\neq a_0.$$

当原假设 H_0 成立时，由(8.2.1)式知，建立了统计量

$$U=\frac{\bar{\xi}-a_0}{\sigma_0/\sqrt{n}},$$

U 服从正态 $N(0,\ 1)$，它可作为对 H_0 的显著性检验的检验统计量. 给定显著性水平 α，可选临界值 u_α，使得

$$P\{|U|>u_\alpha\}=\alpha,$$

即

$$P\left\{-u_\alpha\leqslant\frac{\bar{\xi}-a_0}{\sigma_0/\sqrt{n}}\leqslant u_\alpha\right\}=1-\alpha, \tag{8.4.3}$$

记

$$B_\alpha=(-\infty, -u_\alpha)\cup(u_\alpha, \infty), \tag{8.4.4}$$

则 B_α 为 H_0 的显著性水平为 α 的否定域. (8.4.3)式等价于下式：

$$P\{na_0-\sqrt{n}\,\sigma_0 u_\alpha\leqslant\sum_{i=1}^{n}\xi_i\leqslant na_0+\sqrt{n}\,\sigma_0 u_\alpha\}$$

$$=1-\alpha. \tag{8.4.5}$$

记

$$B_0=(-\infty, na_0-\sqrt{n}\sigma_0 u_\alpha)\cup(na_0+\sqrt{n}\sigma_0 u_\alpha, \infty), \tag{8.4.6}$$

B_0 也是 H_0 的显著性水平为 α 的否定域. 但要注意，检验统计量不再是服从正态 $N(0,1)$ 的U了，而是 $\zeta=\sum_{i=1}^{n}\xi_i$，检验统计量 ζ 服从正态 $N(na_0, \sqrt{n}\sigma_0)$.

以上(8.4.1)、(8.4.2)及(8.4.4)、(8.4.6)等式，都是用检验统计量将其取值空间划分两个不相交的部分，即 H_0 的否定域与不否定域.

从(8.4.6)式看到，当子样 $\xi_1, \cdots, \xi_n$ 的观察值 $(x_1, \cdots, x_n)$ 属于集合

$$\mathscr{X}_0=\left\{(x_1, \cdots, x_n): \sum_{i=1}^{n}x_i \in B_0\right\}, \tag{8.4.7}$$

时，则在显著性水平 α 下否定 H_0. 因此，(8.4.7) 式建立的区域 $\mathscr{X}_0$，也是 H_0 的显著性水平为 α 的否定域. (8.4.7)式建立的 $\mathscr{X}_0$，实际上是将子样空间$\mathscr{X}$剖分为互不相交的两部分，即

$$\mathscr{X}=\mathscr{X}_0\cup(\mathscr{X}-\mathscr{X}_0),$$

$$\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}_0)=\varnothing.$$

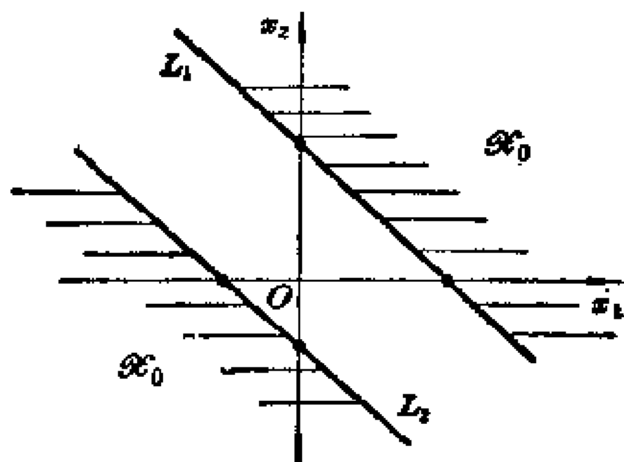


图 8.4.2

其中 $\mathscr{X}_0$ 为 H_0 的显著性水平为 α 的否定域.

我们就 $n=2$ 的简单情形，对否定域 $\mathscr{X}_0$ 作几何解释. 在平面

上作两条平行的直线 L_1 及 L_2

$$L_1:\ x_1+x_2=2a_0+\sqrt{2}\,\sigma_0u(\alpha);$$

$$L_2:\ x_1+x_2=2a_0-\sqrt{2}\,\sigma_0u(\alpha).$$

将子样空间 $\mathscr{X}$ 剖分为互不相交的两部分 $\mathscr{X}=\mathscr{X}_0\cup(\mathscr{X}-\mathscr{X}_0)$，若子样观察值 $(x_1,x_2)\in\mathscr{X}_0$，则否定 H_0，当子样观察值 $(x_1,x_2)\overline{\in}\mathscr{X}_0$ 时，则不否定 H_0.

由此可见，建立了 H_0 的检验统计量后，给定显著性水平 α 时，选择临界值及确定 H_0 的否定域，等价于将子样空间剖分为互不相交的两部分，确定其中的一部分 $\mathscr{X}_0$ 作为 H_0 的否定域. 也就是说，怎样确定 H_0 的否定域，等价于怎样剖分子样空间为互不相交的两部分.

综上所述，在对 H_0 作显著性检验中，给定显著性水平 α 后，可以有不同的临界值，相应地有不同的否定域，这等价于将子样空间剖分为互不相交的两部分的不同剖分法，一个否定域对应于一种剖分法. 因此，在 §8.2 中讨论到的 H_0 的显著性检验法还很不完善.

现在，我们从例 8.4.2 就 $n=2$ 的讨论中，引进一些新的概念，探讨 H_0 的最佳检验.

定义 8.4.1(两类错误) 对于原假设 H_0，设 $\mathscr{X}_0$ 为其否定域. 当 H_0 成立时，若子样观察值 $(x_1,\cdots,x_n)\in\mathscr{X}_0$，则作出否定 H_0 的判断，这时则称作犯第一类错误（弃真）. 犯第一类错误的概率记为

$$P\{(\xi_1,\cdots,\xi_n)\in\mathscr{X}_0\,|\,\Omega_0\}=\alpha. \tag{8.4.8}$$

当 H_0 不成立时，若子样观察值 $(x_1,\cdots,x_n)\overline{\in}\mathscr{X}_0$，而作出不否定 H_0 的判断，这时则称作犯第二类错误(取伪). 犯第二类错误的概率记为

$$P\{(\xi_1,\cdots,\xi_n)\overline{\in}\mathscr{X}_0\,|\,\Omega_1\}=\beta, \tag{8.4.9}$$

其中 $\Omega_1=\Omega-\Omega_0$，Ω为参数空间．如 a 为参数，$\Omega_0=\{a=0\}$，$\Omega_1=\{a\neq 0\}$．对 H_0 的判断情况用表 8.4.1 作说明．

表 8.4.1

假设 / 决策	H_0为真	H_0不真
否定H_0	第一类错误	正确
不否定H_0	正确	第二类错误

为了较直观地理解定义 8.4.1，我们仅就 $a_0<a_1$ 的一种情形，作图说明如下：

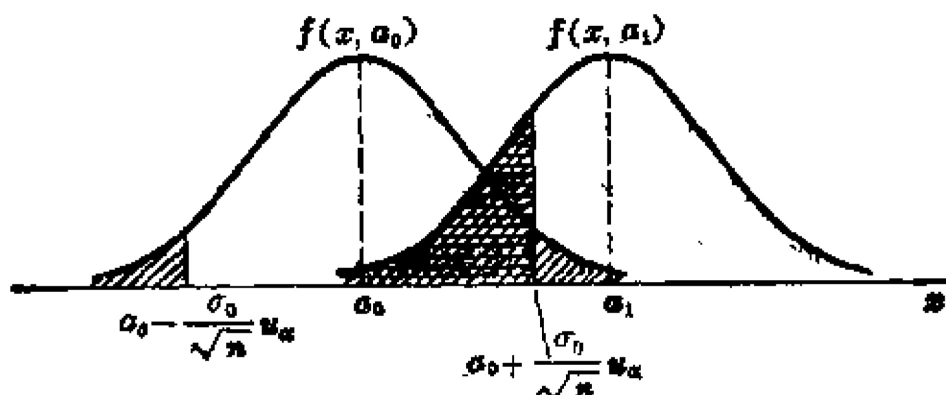


图 8.4.3

图中以斜线部分表示α 的大小，以网格的部分表示β 的大小，其中

$$B_0=\left(-\infty,\ a_0-\frac{\sigma_0}{\sqrt{n}}u_\alpha\right)\cup\left(a_0+\frac{\sigma_0}{\sqrt{n}}u_\alpha,\ \infty\right)$$

$$\mathscr{X}_0=\{(x_1,\cdots,x_n):\ \bar{x}\in B_0\}$$

$$P\{(\xi_1,\cdots,\xi_n)\in\mathscr{X}_0\,|\,a=a_0\}=\alpha;$$

$$P\{(\xi_1,\cdots,\xi_n)\overline{\in}\mathscr{X}_0\,|\,a=a_1\}=\beta.$$

注意，$\beta\neq 1-\alpha$，当 n 固定时，α 减小等价于 u_α 增大，从而等价于β 增大．当α 固定时，则增大 n 等价于减小 β．

在实际工作中这两类错误造成的影响常常是不一样的，在不同的场合允许α及β的大小也常常不一样．例如在检查病人是否患有某种疾病时，无病当作有病与有病当作无病，其造成的后果就不一样．又如产品检查验收中，危及健康的药品验收，希望不合格而当作合格验收的可能性尽量小些，即犯第二类错误的概率β要求很小．而对于价格高昂的商品，生产者希望验收者犯第一类错误的可能性尽量小些，即犯第一类错误的概率α要求很小．

当然，我们总希望所作出的检验能使犯两类错误的概率同时尽可能地小，最好是全为零．就是说，H_0为真时，能作出不否定H_0的判断；而当H_0不真时，能作出否定H_0的判断．这种理想化的检验法，由于子样$\xi_1,\cdots,\xi_n$的随机性，在实际上是不存在的．当子样的容量n给定后，犯这两类错误的概率难于同时被控制．在假设检验的一般理论中，是在这样原则下进行讨论的，即固定一个α（或β）而使得另一个β（或α）尽可能地小．通常是限制犯第一类错误的概率α，即选定一个数α，$0<\alpha<1$，要求犯第二类错误的概率β尽可能地小．当然尽量增大n，可使α及β都尽可能地小．

现将参数的假设检验问题提得更一般化，叙述得更明确些．设总体ξ的分布函数$F(x;\theta)$中含有未知参数θ，参数空间记作Ω，即$\theta\in\Omega$．考虑如下假设检验问题：

$$H_0:\ \theta\in\Omega_0,\quad H_1:\ \theta\in\Omega-\Omega_0.$$

将子样空间$\mathscr{X}$剖分为互不相交的两部分$\mathscr{X}_0$及$\mathscr{X}-\mathscr{X}_0$，引进记号$M(\mathscr{X}_0,\theta)$：对于每一$\theta\in\Omega$，令：

$$M(\mathscr{X}_0,\theta)=P_\theta\{\mathscr{X}_0\}=P\{(\xi_1,\cdots,\xi_n)\in\mathscr{X}_0|\theta\}. \tag{8.4.10}$$

由(8.4.10)式知，

$$M(\mathscr{X}_0,\theta)=\begin{cases}\alpha(\theta), & \theta\in\Omega_0,\\ 1-\beta(\theta), & \theta\in\Omega-\Omega_0.\end{cases} \tag{8.4.11}$$

即当H_0为真时，否定H_0的概率为$\alpha(\theta)$，当H_0不真时，否定H_0

的概率为 $1-\beta(\theta)$. $M(\mathscr{X}_0,\theta)$是否定 H_0 的概率.

定义 8.4.2 势函数（或效函数） 设 $\mathscr{X}_0$ 为原假设 $H_0(\theta\in\Omega_0)$的否定域，称由(8.4.10)定义的函数 $M(\mathscr{X}_0,\theta)$为具有否定域 $\mathscr{X}_0$ 的势函数(或效函数). 若给定常数 $\alpha, 0<\alpha<1$, 否定域 $\mathscr{X}_0$ 满足下述不等式:

$$M(\mathscr{X}_0,\theta)=P_\theta\{\mathscr{X}_0\}\leqslant\alpha \quad (\theta\in\Omega_0),$$

则称 α 为否定域 $\mathscr{X}_0$ 的检验水平.

势函数(或效函数)$M(\mathscr{X}_0,\theta)$依赖于 $\mathscr{X}_0$ 及参数 θ, 当选定一个否定域 $\mathscr{X}_0$ 时, 它是参数 θ 的一元函数. 势函数的意义在于当 H_0 不真时, 反映否定 H_0 的功效大小. 因此, 寻找 H_0 的最佳检验法, 可归结为选择这样的否定域 $\mathscr{X}_0$, 当 H_0 不真时, 使得否定 H_0 的功效最大.

如果参数空间 $\Omega=\{\theta_0,\theta_1\}$, $\Omega_0=\{\theta_0\}$, $\Omega_1=\{\theta_1\}$, 那么势函数为

$$M(\mathscr{X}_0,\theta)=\begin{cases}\alpha, & \theta=\theta_0,\\ 1-\beta, & \theta=\theta_1.\end{cases} \tag{8.4.12}$$

例 8.4.3 设总体 ξ 服从正态 $N(a,\sigma)$, σ 为已知, 记作 σ_0. 考虑简单原假设及简单备选假设

$$H_0:\ a=0, \quad H_1:\ a=a_1(\neq 0).$$

试求 u 检验法的势函数及犯第二类错误的概率.

解 对于显著性水平 $\alpha=0.05$, 当原假设 H_0 成立时, 由(8.4.3)式知道, u 检验法的否定域为

$$\mathscr{X}_0=\left\{(x_1,\cdots,x_n): \frac{|\bar{x}|}{\sigma_0/\sqrt{n}}\geqslant 1.96\right\}$$

$$=\left\{(x_1,\cdots,x_n): |\bar{x}|\geqslant\frac{1.96\sigma_0}{\sqrt{n}}\right\},$$

对应的势函数为

$$M(\mathscr{X}_0, a)=P_a\{\mathscr{X}_0\}=P_a\left\{\frac{|\bar{\xi}|}{\sigma_0}\sqrt{n}\geqslant 1.96\right\}$$

$$=P_a\left\{\frac{\bar{\xi}}{\sigma_0}\sqrt{n}\leqslant -1.96\right\}+P_a\left\{\frac{\bar{\xi}}{\sigma_0}\sqrt{n}\geqslant 1.96\right\}$$

$$=P_a\left\{\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}}\leqslant -1.96-\frac{a}{\sigma_0/\sqrt{n}}\right\}$$

$$+P_a\left\{\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}}\geqslant 1.96-\frac{a}{\sigma_0/\sqrt{n}}\right\}$$

$$=1-\Phi\left(1.96-\frac{a}{\sigma_0/\sqrt{n}}\right)+\Phi\left(-1.96-\frac{a}{\sigma_0/\sqrt{n}}\right),$$

其中 $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$. 当 H_0 为真时，即 $a=0$ 时，则

$$M(\mathscr{X}_0, 0)=1-\Phi(1.96)+\Phi(-1.96)=0.05,$$

当 H_1 为真时，即 $a=a_1(\neq 0)$ 时，则有

$$M(\mathscr{X}_0, a_1)=1-\Phi\left(1.96-\frac{a_1}{\sigma_0/\sqrt{n}}\right)$$

$$+\Phi\left(-1.96-\frac{a_1}{\sigma_0/\sqrt{n}}\right),$$

即得

$$\beta=1-M(\mathscr{X}_0, a_1)$$

$$=\Phi\left(1.96-\frac{a_1}{\sigma_0}\sqrt{n}\right)-\Phi\left(-1.96-\frac{a_1}{\sigma_0}\sqrt{n}\right).$$

对于一般显著性水平 $0<\alpha<1$, u 检验的势函数为

$$M(\mathscr{X}_0, a)=1-\Phi\left(u_\alpha-\frac{a}{\sigma_0}\sqrt{n}\right)+\Phi\left(-u_\alpha-\frac{a}{\sigma_0}\sqrt{n}\right),$$

它与 α、a、n 都有关. 记 $\delta=\frac{a}{\sigma_0}\sqrt{n}$.

为了便于研究势函数的有关性质，不妨设 $a>0$，即有 $\delta>0$. 于是

$$M(\mathscr{X}_0, a) = 1 - \Phi(u_\alpha - \delta) + \Phi(-u_\alpha - \delta).$$

将其分别对 u_α 及 δ 求导，则有：

$$\frac{\partial M}{\partial \delta} = \Phi'(u_\alpha - \delta) - \Phi'(-u_\alpha - \delta)$$

$$= \Phi'(u_\alpha - \delta) - \Phi'(u_\alpha + \delta) > 0,$$

$$\frac{\partial M}{\partial u_\alpha} = -\Phi'(u_\alpha - \delta) - \Phi'(-u_\alpha - \delta)$$

$$= -\Phi'(u_\alpha - \delta) - \Phi'(u_\alpha + \delta) < 0.$$

这表明当 α 及 a 固定时，$M(\mathscr{X}_0, a)$ 是 n 的严格增函数；当 n 及 a 固定时，$M(\mathscr{X}_0, a)$ 是 u_α 的严格减函数，也就是 α 的严格减函数. 当 n 及 α 固定时，则 $M(\mathscr{X}_0, a)$ 是 $a(>0)$ 的严格增函数，并且可由

$$\frac{\partial^2 M}{\partial \delta^2} = -\Phi''(u_\alpha - \delta) - \Phi''(n_\alpha + \delta)$$

判断它的凹凸性. 由此可作出势函数曲线图.

在图 8.4.4 中，对各种不同的 n，在 $\alpha = 0.05$ 及 $a > 0$ 的情形，给出了该势函数 $M(\mathscr{X}_0, a)$ 的具体曲线形式. 从图上看到，子样容

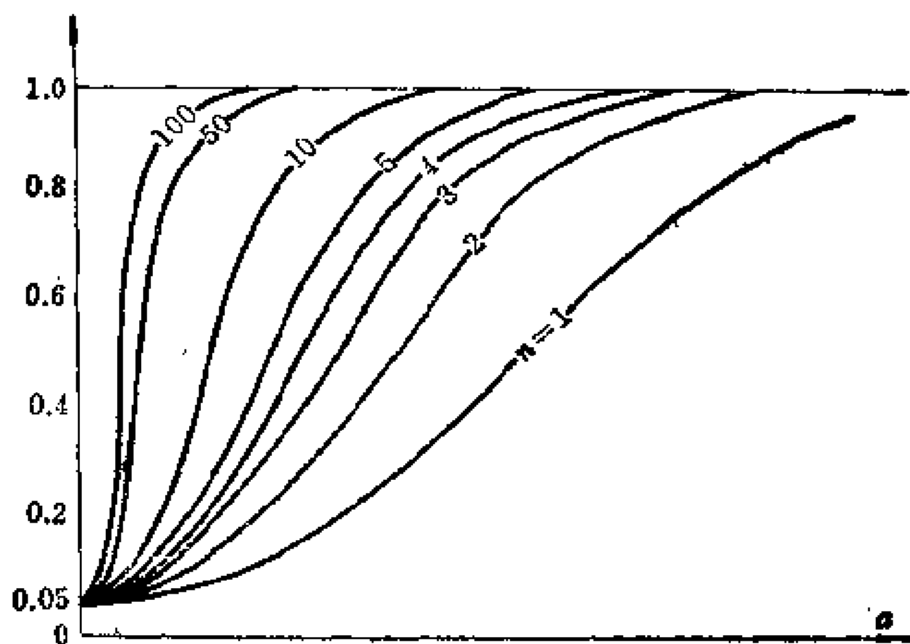


图 8.4.4

量 n 越大，势函数越陡，对应的检验的功效越大.

定义 8.4.3(最优势检验) 设参数空间为 Ω，考虑

$$H_0:\ \theta_0\in\Omega_0,\quad H_1:\ \theta_1\in\Omega_1=\Omega-\Omega_0,$$

称具有检验水平 α 的否定域 $\mathscr{X}_0$ 所确定的检验为最优势检验（简称作最佳检验)，是指对于具有检验水平 α 的任一否定域 $\mathscr{X}_0'$，对每一 $\theta_1\in\Omega_1$，有

$$\begin{aligned}P_{\theta_1}\{\mathscr{X}_0\}&=M(\mathscr{X}_0,\theta_1)\\&\geqslant M(\mathscr{X}_0',\theta_1)=P_{\theta_1}\{\mathscr{X}_0'\}\end{aligned}$$

其中

$$\begin{aligned}&M(\mathscr{X}_0,\theta_0)\leqslant\alpha,\quad \theta_0\in\Omega_0,\\&M(\mathscr{X}_0',\theta_0)\leqslant\alpha,\quad \theta_0\in\Omega_0.\end{aligned}$$

奈曼-皮尔逊定理 设总体 ξ 的分布函数 $F(x;\theta)$ 中含未知参数 θ，$\theta\in\Omega=\{\theta_0,\theta_1\}$，则对任一数 α，$0<\alpha<1$，对于假设检验问题：

$$H_0:\theta=\theta_0,\quad H_1:\theta=\theta_1.$$

必存在水平为 α 的最佳检验.

从定义 8.4.3 看到，寻找 H_0 的最佳检验，归纳为寻找 H_0 的最佳否定域.

奈曼-皮尔逊定理指出，对于原假设及备选假设都是简单的情形，最佳检验是存在的，并给出了用似然比法确定最佳否定域.

如果备选假设是复合的，最佳检验不一定存在. 如果存在，称它为一致最佳检验. 应用似然比法确定的最佳否定域，如果它与备选假设中的 θ_1 无关，那么它就是一致最佳否定域，也即为一致最佳检验.

奈曼-皮尔逊定理对于连续型分布及离散型分布其结论都同样成立. 下面叙述连续型分布的情形. 在叙述过程中，如用概率函数 $P(x,\theta)$ 代替密度函数 $f(x,\theta)$ 的位置，用求和符号代替积分符号，就成为对离散型分布的叙述.

在证明奈曼-皮尔逊定理之前，先引进似然比的概念，并证明一个引理.

若总体 ξ 有密度函数 $f(x,\theta)$，$\xi_1,\cdots,\xi_n$ 为其子样，θ 为未知参数，$\theta\in\Omega$. 在第七章的 §7.1 中引进了似然函数的概念，即

$$L(\theta)=\prod_{i=1}^{n} f(\xi_i;\theta)$$

为未知参数 θ 的似然函数. 现称

$$\frac{L(\theta_1)}{L(\theta_0)}=\frac{\prod\limits_{i=1}^{n} f(\xi_i,\theta_1)}{\prod\limits_{i=1}^{n} f(\xi_i,\theta_0)} \tag{8.4.13}$$

为似然比. 用似然比将子样空间 $\mathscr{X}$ 剖分为互不相交的两部分，其中

$$\mathscr{X}_0=\{x_1,\cdots,x_n):L(\theta_1)\geqslant cL(\theta_0)\}, \tag{8.4.14}$$

$$\mathscr{X}-\mathscr{X}_0=\{x_1,\cdots,x_n):L(\theta_1)<cL(\theta_0)\},$$

其中 c 为非负的实数. 定义实数 c 的函数 $\varphi(c)$ 为

$$\varphi(c)=P_{\theta_0}\{\mathscr{X}_0\}=P_{\theta_0}\left\{\frac{L(\theta_1)}{L(\theta_0)}\geqslant c\right\}. \tag{8.4.15}$$

*引理　由(8.4.15)所定义的函数 $\varphi(c)$ $(c\geqslant 0)$，为一左连续、单调不增的函数，而且

$$\varphi(0)=1,\ 0\leqslant\varphi(c)\leqslant\frac{1}{c},$$

$$\lim_{c\to\infty}\varphi(c)=0.$$

由此结论可见 $1-\varphi(c)$ 为一分布函数.

并且对任一 α，$0<\alpha<1$，存在常数 c_α，使得

$$\varphi(c_\alpha+0)\leqslant\alpha\leqslant\varphi(c_\alpha)$$

其中

$$\varphi(c_\alpha+0)=P_{\theta_0}\left\{\frac{L(\theta_1)}{L(\theta_0)}>c_\alpha\right\}$$

证 由似然函数的非负性及 (8.4.15) 式的定义知道，$\varphi(0)=1, 0\leqslant\varphi(c)\leqslant 1$. 而对于 $c>0$, 有

$$\varphi(c)=\int_{\mathscr{X}_0}\cdots\int\prod_{i=1}^{n}f(x_i,\theta_0)\,dx_1\cdots dx_n$$

$$\leqslant\int_{\mathscr{X}_0}\cdots\int\frac{1}{c}\prod_{i=1}^{n}f(x_i,\theta_1)\,dx_1\cdots dx_n$$

$$\leqslant\frac{1}{c},$$

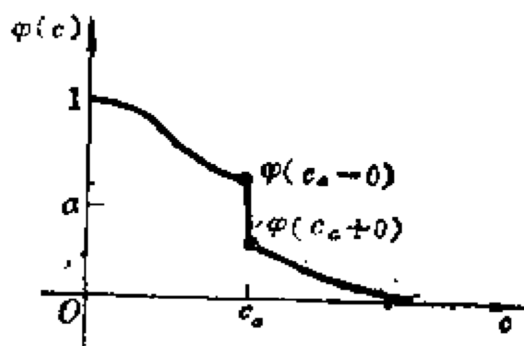


图 8.4.5

因而

$$\lim_{c\to\infty}\varphi(c)=0.$$

又由(8.4.15)定义知，若 $0\leqslant c_1<c_2$,
则

$$\varphi(c_1)\geqslant\varphi(c_2).$$

还可证明当 $c_n\uparrow c$ 时，则有 $\lim\limits_{n\to\infty}\varphi(c_n)=\varphi(c)$

函数 $\varphi(c)$ 的图形如图 8.4.5 所示.

***奈曼-皮尔逊定理的证明**

(1) 若存在常数 c_α, 使得 $\varphi(c_\alpha)=\alpha$, 则检验水平为 α 的最佳否定域为

$$\mathscr{X}_0=\{(x_1,\cdots,x_n): L(\theta_1)\geqslant c_\alpha L(\theta_0)\}. \qquad (8.4.16)$$

证 设 $\mathscr{X}_0'$ 是检验水平为 α 的任一否定域，即

$$P_{\theta_0}\{\mathscr{X}_0\}=P_{\theta_0}\{\mathscr{X}_0'\}=\alpha.$$

记

$$\mathscr{X}_0=[\mathscr{X}_0\cap\mathscr{X}_0']\cup[\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}_0')],$$

$$\mathscr{X}_0'=[\mathscr{X}_0'\cap\mathscr{X}_0]\cup[\mathscr{X}_0'\cap(\mathscr{X}-\mathscr{X}_0)].$$

易见

$$P_{\theta_1}\{\mathscr{X}_0\}-P_{\theta_1}\{\mathscr{X}_0'\}=\underset{\mathscr{X}_0}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i]$$

$$-\underset{\mathscr{X}_0'}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i]$$

$$=\underset{\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}_0')}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i]$$

$$-\underset{\mathscr{X}_0'\cap(\mathscr{X}-\mathscr{X}_0)}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i].$$

由(8.4.16)式知，若对于子样的每一次观察值$(x_1,\cdots,x_n)\in\mathscr{X}_0$，则有

$$\prod_{i=1}^{n}f(x_i,\theta_1)\geqslant c_\alpha\prod_{i=1}^{n}f(x_i,\theta_0).$$

因此

$$\underset{\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}_0')}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i]$$

$$\geqslant c_\alpha\underset{\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}_0')}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i)$$

$$\underset{\mathscr{X}_0'\cap(\mathscr{X}-\mathscr{X}_0)}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_1)dx_i],$$

$$<c_\alpha\underset{\mathscr{X}_0'\cap(\mathscr{X}-\mathscr{X}_0)}{\int\cdots\int}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i],$$

于是

$$P_{\theta_1}\{\mathscr{X}_0\}-P_{\theta_1}\{\mathscr{X}_0'\}$$

$$\geqslant c_\alpha\left\{\int\cdots\int_{\mathscr{X}_0\cap(\mathscr{X}-\mathscr{X}'_0)}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i]\right.$$

$$\left.-\int\cdots\int_{\mathscr{X}'_0\cap(\mathscr{X}-\mathscr{X}_0)}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i]\right\}$$

$$=c_\alpha\left\{\int\cdots\int_{\mathscr{X}_0}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i]\right.$$

$$\left.-\int\cdots\int_{\mathscr{X}'_0}\prod_{i=1}^{n}[f(x_i,\theta_0)dx_i]\right\}$$

$$=c_\alpha\{\alpha-\alpha\}=0.$$

因此，

$$P_{\theta_1}\{\mathscr{X}'_0\}\leqslant P_{\theta_1}\{\mathscr{X}_0\}$$

定理得证.

(2) 若存在常数 c_α，满足

$$\varphi(c_\alpha+0)=\alpha\leqslant\varphi(c_\alpha),$$

则检验水平 α 的最佳否定域为

$$\mathscr{X}_0^*=\{(x_1,\cdots,x_n):\ L(\theta_1)>c_\alpha L(\theta_0)\}. \qquad (8.4.17)$$

证明方法与(1)相同.

(3) 若为 $\varphi(c_\alpha+0)<\alpha<\varphi(c_\alpha)$，则检验水平 α 的最佳否定域为

$$\mathscr{X}_0^{**}=\mathscr{X}_0^*\cup B_0^*, \qquad (8.4.18)$$

其中 $\mathscr{X}_0^*$ 由(8.4.17)式给出，$B_0^*\subset B_0^\gamma\subset B_0$，都为 n 维空间中的子集，满足条件

$$P_{\theta_0}\{B_0^*\}=\sup_\gamma[P_{\theta_0}\{B_0^\gamma\}<\alpha-\varphi(c_\alpha+0)]$$

$$\leqslant\alpha-\varphi(c_\alpha+0),$$

其中

$$B_0=\{(x_1,\cdots,x_n):\ L(\theta_1)=c_\alpha L(\theta_0)\},$$

$$B_0^r=\{(-\infty,r_1)\times\cdots\times(-\infty,r_n)\}\cap B_0.$$

证 令

$$g(r)=P_{\theta_0}\{B_0^r\},\tag{8.4.19}$$

则 $g(r)$ 具有如下性质：对任一 $r^*=(r_1^*,\cdots,r_n^*)$，

$$\lim_{\substack{r_i\to r_i^*\\ 1\leqslant i\leqslant n}} g(r)=g(r^*).$$

又令

$$\alpha^*=\sup_r\{g(r):\ g(r)<\alpha-\varphi(c_\alpha+0)\}$$

则依上确界定义知，存在 $r^m\nearrow$（是指 n 个分量都递增），使得

$$\alpha^*-\frac{1}{m}<g(r^m)\leqslant\alpha^*.$$

于是

$$\lim_{m\to\infty}g(r^m)=\alpha^*.$$

另一方面，由单调性知必存在某个 r^*，使得 $r^m\nearrow r^*$，从而

$$\alpha^*=\lim_{m\to\infty}g(r^m)=g(r^*)\leqslant\alpha-\varphi(c_\alpha+0).$$

最后，令

$$B_0^*=B_0^{r^*}=B_0\cap\{(-\infty,r_1^*)\times\cdots\times(-\infty,r_n^*)\},$$

则 B_0^* 为所求，也就是说，$\mathscr{X}_\theta^*$ 是检验水平 α 的否定域，即有

$$P_{\theta_0}\{\mathscr{X}_0^{**}\}=P_{\theta_0}\{\mathscr{X}_0^*\}+P_{\theta_0}\{B_0^*\}\leqslant\alpha.$$

例 8.4.4 设总体 ξ 服从正态 $N(a,\sigma_0)$，其中 σ_0 为已知，考虑下列各种情形的假设检验问题：

(A) H_0: $a=a_0$ H_1: $a=a_1(>a_0)$,

(B) H_0: $a=a_0$ H_1: $a=a_1(<a_0)$,

(C) H_0: $a=a_0$ H_1: $a>a_0$,

(D) H_0: $a\leqslant a_0$ H_1: $a>a_0$,

(E) H_0: $a\geqslant a_0$ H_1: $a<a_0$.

给定显著性水平 α, $0<\alpha<1$, 试求最佳否定域.

解 由(8.4.13)式知,似然比为

$$\frac{L[a_1]}{L[a_0]}=\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}[(\xi_i-a_1)^2-(\xi_i-a_0)^2]\right\}$$

$$=\exp\left\{\frac{n(a_1-a_0)\bar{\xi}}{\sigma_0^2}-\frac{n(a_1^2-a_0^2)}{2\sigma_0^2}\right\}$$

$$=\exp\left\{\frac{\sqrt{n}(a_1-a_0)}{\sigma_0}\left[\frac{\sqrt{n}(\bar{\xi}-a_0)}{\sigma_0}-\frac{\sqrt{n}(a_1-a_0)}{2\sigma_0}\right]\right\}$$

$$=e^{kU-\frac{1}{2}k_2} \tag{8.4.20}$$

其中

$$U=\sqrt{n}\frac{\bar{\xi}-a_0}{\sigma_0},\qquad k=\sqrt{n}\frac{a_1-a_0}{\sigma_0}.$$

$$\varphi(c)=P_{a_0}\{(\xi_1,\cdots,\xi_n):\ e^{kU-\frac{1}{2}k_2}\geqslant c\}$$

k 为常数,U 服从正态分布. $k_2=n(a_1^2-a_0^2)/\sigma_0^2$.

现对于给定的显著性水平 α, $0<\alpha<1$, 讨论(A)及(B)的情形,确定最佳否定域.

(A):由

$$\int_b^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\alpha$$

可确定 b 值,记 $b=u_\alpha$, 选取 $c_\alpha=e^{ku_\alpha-\frac{1}{2}k_2}$ 时,则

$$\mathscr{X}_0=\{(x_1,\cdots,x_n):L(a_1)\geqslant c_\alpha L(a_0)\}$$

为 H_0 的最佳否定域,事实上,由于 $k>0$, 则

$$\mathscr{X}_0=\{e^{kU-\frac{1}{2}k_2}\geqslant e^{ku_\alpha-\frac{1}{2}k_2}\}=\{U\geqslant u_\alpha\}. \tag{8.4.21}$$

当 H_0 为真时,则 U 服从 $N(0,1)$, 于是

$$\varphi(c_\alpha)=P_{a_0}\{\mathscr{X}_0\}=P_{a_0}\{U\geqslant u_\alpha\}=\alpha.$$

(B)：由

$$\int_{-\infty}^{b}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\alpha$$

可确定 b 值，记 $b=-u_\alpha$. 选取 $c_\alpha=e^{-ku_\alpha-\frac{1}{2}k^2}$ 时，则

$$\mathscr{X}_0=\{(x_1,\cdots,x_n):L(a_1)\geqslant c_\alpha L(a_0)\}$$

为 H_0 的最佳否定域. 事实上，由于 $k<0$，则

$$\mathscr{X}_0=\{e^{kU-\frac{1}{2}k^2}\geqslant e^{-ku_\alpha-\frac{1}{2}k^2}\}=\{U\leqslant -u_\alpha\}. \qquad (8.4.22)$$

当 H_0 为真时，则 U 服从正态 $N(0,1)$，于是

$$\varphi(c_\alpha)=P_{a_0}\{\mathscr{X}_0\}=P_{a_0}\{U\leqslant -u_\alpha\}=\alpha.$$

下面，利用(A)及(B)的结论，试求(C)，(D)，(E)的一致最佳否定域.

(C)：对于 H_1，任取 a_1，满足 $a_1>a_0$，由(A)知，H_0: $a=a_0$，H_1': $a=a_1(>a_0)$，最佳否定域由(8.4.21)式确定. 由于这个最佳否定域与 a_1 具体值无关，只要求 $a_1>a_0$，因而它为(C)的一致最佳否定域.

(D)：对于 H_0': $a=a_0$，H_1: $a>a_0$ 而言，由(C)知，检验水平为 α 的一致最佳否定域为

$$\mathscr{X}_0=\{(x_1,\cdots,x_n):U\geqslant u_\alpha\}.$$

可以证明 $\mathscr{X}_0$ 也是(D)的一致最佳否定域. 事实上，对任何 $a\leqslant a_0$，有

$$\begin{aligned}P_a\{\mathscr{X}_0\}&=P\left\{\frac{\bar{\xi}-a_0}{\sigma_0}\sqrt{n}\geqslant u_\alpha\,|\,a\right\}\\&=P\left\{\frac{\bar{\xi}-a}{\sigma_0}\sqrt{n}\geqslant u_\alpha+\frac{a_0-a}{\sigma_0}\sqrt{n}\,|\,a\right\}\\&\leqslant P\left\{\frac{\bar{\xi}-a}{\sigma_0}\sqrt{n}\geqslant u_\alpha\,|\,a\right\}=\alpha,\end{aligned}$$

$$\sup_{a\leqslant a_0}[P_a\{\mathscr{X}_0\}]=P_{a_0}\{\mathscr{X}_0\}=\alpha.$$

设 $\mathscr{X}_0'$ 是水平为 α 的任一否定域，即对每一 $a\leqslant a_0$，

$$P_a\{\mathscr{X}_0'\}\leqslant\alpha,$$

特别地有

$$P_{a_0}\{\mathscr{X}_0'\}\leqslant\alpha,$$

即 $\mathscr{X}_0'$ 为 H_0'、H_1 的水平为 α 的另一否定域．由 (C)知，对任何 $a>a_0$，有

$$P_a\{\mathscr{X}_0'\}\leqslant P_a\{\mathscr{X}_0\},$$

从而 $\mathscr{X}_0$ 为(D)的水平为 α 的一致最佳否定域．

(E) 按(C)及(D)的证明方法可知，(8.4.22)式的 $\mathscr{X}_0$ 为(E)的水平为 α 的一致最佳否定域．

例 8.4.5 设总体 ξ 服从两点分布，其概率分布为

$$P\{x;p\}=\begin{cases}p\ , & x=1,\\ 1-p, & x=0,\end{cases}$$

其中 $0<p<1$，$\xi_1,\cdots,\xi_n$ 为其子样．考虑假设检验问题：

$$H_0: p=p_0,\quad H_1: p=p_1$$

$(p_1>p_0)$．试求最佳检验即最佳否定域．

解 由(8.4.13)式知，似然比为

$$\frac{L(p_1)}{L(p_0)}=\frac{\prod\limits_{i=1}^{n}P(\xi_i,p_1)}{\prod\limits_{i=1}^{n}P(\xi_i,p_0)}=\frac{\prod\limits_{i=1}^{n}\left[p_1^{\xi_i}\ (1-p_1)^{1-\xi_i}\right]}{\prod\limits_{i=1}^{n}\left[p_0^{\xi_i}\ (1-p_0)^{1-\xi_i}\right]}$$

$$=\left(\frac{p_1}{p_0}\right)^{\sum\limits_{i=1}^{n}\xi_i}\left(\frac{1-p_1}{1-p_0}\right)^{n-\sum\limits_{i=1}^{n}\xi_i}.$$

记

$$\zeta=\sum_{i=1}^{n}\xi_i,$$

由 $p_1>p_0$ 知，$p_1(1-p_0)>p_0(1-p_1)$，因此

$$\frac{L(p_1)}{L(p_0)}=\left(\frac{p_1}{p_0}\right)^{\zeta}\left(\frac{1-p_1}{1-p_0}\right)^{n-\zeta}$$

$$=\left(\frac{1-p_1}{1-p_0}\right)^{n}\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right]^{\zeta}\geqslant c,$$

它等价于

$$\zeta\geqslant\frac{n\log\left[\dfrac{c^{\frac{1}{n}}(1-p_0)}{(1-p_1)}\right]}{\log\left[\dfrac{p_1(1-p_0)}{p_0(1-p_1)}\right]}.$$

如果存在常数 c_α，使得

$$\varphi(c_\alpha)=P_{p_0}\{\mathscr{X}_0\}=\alpha,$$

其中

$$\mathscr{X}_0=\{(x_1,\cdots,x_n):L(p_1)\geqslant c_\alpha L(p_0)\},$$

这等价于存在一个数 b，满足

$$P_{p_0}\{\zeta\geqslant b\}=\alpha,$$

其中 $b=n\log\left[c_\alpha^{\frac{1}{n}}\left(\dfrac{1-p_0}{1-p_1}\right)\right]\Big/\log\left[\dfrac{p_1(1-p_0)}{p_0(1-p_1)}\right]$，而统计量 ζ 服从二项分布，即有

$$P\{\zeta=k\}=C_n^k p^k(1-p)^{n-k}\qquad(k=0,1,\cdots,n).$$

现举例说明：$n=5,\alpha=0.05,p_0=0.5,p_1=0.6$，则有

$$P_{p_0}\{\zeta=5\}=\frac{1}{32};\qquad P_{p_0}\{\zeta=4\}=\frac{5}{32}.$$

于是

$$P_{p_0}\{\zeta\geqslant5\}=\frac{1}{32}<0.05,$$

$$P_{p_0}\{\zeta\geqslant4\}=\frac{6}{32}>0.05.$$

即有

$$P_{p_0}\{\zeta>4\}=\frac{1}{32}<0.05<P_{p_0}\{\zeta\geqslant4\}=\frac{6}{32}$$

选取 $C_\alpha=4$，则水平为 $\alpha=0.05$ 的最佳否定域为

$$\mathscr{X}_0^{**}=\mathscr{X}_0^*=\{\zeta_n>4\}$$

因为形如 $\{\zeta_n=4\}$ 的集合是由形如

$$\{\xi_1=1,\xi_2=1,\xi_3=1,\xi_4=1,\xi_5=0\}$$

的集合构成（即五维实空间中的 5 个点），则它的任一子集，其概率至少为 $\frac{1}{32}$，而

$$\alpha-\varphi(C_\alpha+0)=0.05-\frac{1}{32}=0.01875<\frac{1}{32}$$

于是 B_0^* 为空集.

最后指出，$\mathscr{X}_0=\{\zeta_n>4\}$ 还是检验问题

$$H_0:\ p\leqslant 0.5 \qquad H_1:\ p>0.5$$

的一致最佳检验，从而提供了对于例 8.2.1 求一致最佳检验的方法.

按照例 8.4.4 的证明方法，可以完善上述检验问题的讨论. 对于 H_0: $p=p_0$，H_1: $p>p_0$，先求出 H_0: $p=p_0$，H_1': $p=p_1(>p_0)$ 的最佳否定域，为 $\{\zeta_n>4\}$，它与 p_1 无关，因而它为 H_0、H_1 的一致最佳否定域. 对于 H_0: $p\leqslant p_0$，H_1: $p>p_0$，先讨论 H_0': $p=p_0$，H_1: $p>p_0$ 的最佳检验，继而得出 $\{\zeta_n>4\}$ 也为 H_0: $p\leqslant p_0$，H_1: $p>p_0$ 的一致最佳否定域. 在此 $p_0=0.5$. 对于一般给定的 p_0 值，上述讨论也成立.

*二、无偏检验

我们必须明确，一致最佳检验只在很少情形下存在. 下面不加证明地举出由奈曼发现的定理，它常被用来决定是否有一致最佳检验存在.

奈曼定理 设总体 ξ 的分布函数 $F(x,\theta)$ 中含有未知参数 $\theta\in\Omega$. 对于如下的假设检验问题：

$$H_0:\theta=\theta_0,\ H_1:\theta=\theta_1\quad(\neq\theta_0),$$

如果总体 ξ 的密度函数(或概率函数)对于未知参数 θ 是可微的(即导数存在),那么对于原假设 H_0 没有一致最佳检验.

不存在一致最佳检验时,表明不可能建立这样一个检验法,它相对于原假设的所有可能备选假设,一致地在所有可能的检验中是最佳的. 此时,一个自然的想法是对检验法提出某种合适的要求,然后在满足这种要求的较小的检验类中寻找最优的检验法. 其中一种简单的限制是所谓无偏的要求.

定义 8.4.4 (无偏检验) 对于检验问题:

$$H_0:\theta\in\Omega_0,\ \ H_1:\theta\in\Omega-\Omega_0$$

给定显著性水平 α,如果 H_0 有否定域 $\mathscr{K}_0$,其势函数 $M(\mathscr{K}_0,\theta)$ 满足条件:

$$\begin{aligned}&M(\mathscr{K}_0,\theta)\leqslant\alpha,\ \ \theta\in\Omega_0;\\&M(\mathscr{K}_0,\theta)\geqslant\alpha,\ \ \theta\in\Omega-\Omega_0.\end{aligned}\tag{8.4.23}$$

则称 $\mathscr{K}_0$ 所确定的检验法是具有水平为 α 的无偏检验.

如果存在一个水平为 α 的无偏检验 $\mathscr{K}_0$,它对于任何水平小于或等于 α 的无偏检验 $\mathscr{K}_0'$,对于 $\theta\in\Omega-\Omega_0$ 均满足下述不等式:

$$M(\mathscr{K}_0,\theta)\geqslant M(\mathscr{K}_0',\theta)\tag{8.4.24}$$

则称 $\mathscr{K}_0$ 所确定的检验法是具有水平为 α 的一致最优势无偏检验. 简称作一致最佳无偏检验.

对于一大类检验问题,不存在一致最佳检验,但存在一致最佳无偏检验.

显然,对于无偏检验 $\mathscr{K}_0$,如果它的势函数 $M(\mathscr{K}_0,\theta)$ 是 θ 的连续函数时,则有一子集 $\omega\subset\Omega$,当 $\theta\in\omega$ 时,有

$$M(\mathscr{K}_0,\theta)=\alpha,\tag{8.4.25}$$

其中 ω 是 Ω_0 和 $\Omega-\Omega_0$ 的公共边界. 满足条件(8.4.25)的检验,称为是边界相似的.

定义 8.4.5 相似检验 对于检验问题:

$$H_0:\theta\in\Omega_0,\ \ H_1:\theta\in\Omega-\Omega_0$$

给定显著性水平 α,如果 H_0 有否定域 $\mathscr{K}_0$,其势函数 $M(\mathscr{K}_0,\theta)$ 满足条件

$$M(\mathscr{K}_0,\theta)=\alpha,\ \ \theta\in\Omega_0,\tag{8.4.26}$$

则称 $\mathscr{K}_0$ 所确定的检验法为相似检验. 它是关于 Ω_0 相似的.

一般说来,构造边界相似的检验比构造水平为 α 的无偏检验更加方便,所以下面这个定理在构造一致最佳无偏检验方面有其意义.

定理 设总体 ξ 具有密度函数 $f(x,\theta)$,未知参数 $\theta\in\Omega$,且对于每个检验 $\mathscr{K}_0$,其势函数 $M(\mathscr{K}_0,\theta)$ 是 θ 的连续函数. 如果有一检验 $\mathscr{K}_0^*$,它是所有满

足条件(8.4.25)的检验类中的一致最佳检验，且满足条件

$$M(\mathscr{K}_0^*, \theta) \leqslant \alpha, \quad \theta \in \Omega_0, \tag{8.4.27}$$

则 $\mathscr{K}_0^*$ 是一致最佳无偏检验.

证　要证的是，$\mathscr{K}_0^*$ 满足条件(8.4.23)及(8.4.24)两式. 首先，由于满足条件(8.4.25)式的检验类中包含了无偏检验类，所以由题设知 $\mathscr{K}_0^*$ 比任何一个无偏检验都更优势. 我们来证 $\mathscr{K}_0^*$ 也是一个无偏检验. 由(8.4.27)式知，只需证

$$M(\mathscr{K}_0^*, \theta) \geqslant \alpha, \quad (\theta \in \Omega - \Omega_0),$$

而它是显然的，这是因为 $\mathscr{K}_0^*$ 是满足条件(8.4.25)的检验类中的一致最佳检验. ▌

现在对单参数指数族分布的双边假设检验问题，构造一致最佳无偏检验. 对于如二项分布、负二项分布、泊松分布、数学期望已知或方差已知的正态分布等单参数指数族分布，它们的各类型的单边假设和双边假设检验问题，或者找到了它的一致最佳检验；或者是找到了它们的一致最佳无偏检验.

定理　设总体 ξ 服从单参数指数型的分布，$\xi_1, \cdots, \xi_n$ 为其子样，假定它们的联合密度函数(连续型情形)能表成如下形式

$$\prod_{i=1}^{n} f(x_i; \theta) = R(\theta) \exp\{Q(\theta) T(x_1, \cdots, x_n)\} h(x_1, \cdots, x_n) \tag{8.4.28}$$

(对于离散型情形，用概率函数 $P(x, \theta)$ 代替密度函数 $f(x; \theta)$ 的位置，(8.4.28)式照样成立). 其中 θ 是一个实值参数，$\theta \in \Omega$，$Q(\theta)$ 是 θ 的严格单调上升函数；$h(x_1, \cdots, x_n)$ 及 $T(x_1, \cdots, x_n)$ 只与子样有关. 那么对于

$$H_0: \theta = \theta_0, \qquad H_1: \theta \neq \theta_0,$$

双边假设检验问题及水平 α，存在一致最佳无偏检验，且其否定域 $\mathscr{K}_0$ 的示性函数为

$$\varphi(x_1, \cdots, x_n) = \begin{cases} 1 & T(x_1, \cdots, x_n) < c_1 \text{ 或} > c_2; \\ b_j & T(x_1, \cdots, x_n) = c_j \ (j = 1, 2); \\ 0 & c_1 < T(x_1, \cdots, x_n) < c_2, \end{cases} \tag{8.4.29}$$

其中常数 b_j 及 $c_j \ (j = 1, 2)$，由下式确定：

$$\left.\begin{aligned} &E_{\theta_0}[\varphi(\xi_1, \cdots, \xi_n)] = M(\mathscr{K}_0, \theta_0) = \alpha, \\ &E_{\theta_0}[\varphi(\xi_1, \cdots, \xi_n) T(\xi_1, \cdots, \xi_n)] = \alpha E_{\theta_0}[T(\xi_1, \cdots, \xi_n)]. \end{aligned}\right\} \tag{8.4.30}$$

这个定理的证明，参看[23].

下面，作为定理的直接应用，举两个例子. 对于每一个例子，先是将子样

的联合密度函数或概率函数表成(8.4.28)式，指出检验统计量 $T(\xi_1,\cdots,\xi_n)$ 是什么，然后按(8.4.30)式，对于给定的水平 α，确定 b_1, b_2 及 c_1, c_2, 从而确定(8.4.29)式，找到一致最佳无偏检验.

例 8.4.6 设总体 ξ 服从正态 $N(a_0, \sigma)$, a_0 为已知, σ 为未知参数. $\xi_1,\cdots,\xi_n$ 为其子样，考虑检验问题:

$$H_0:\sigma^2=\sigma_0^2, \qquad H_1:\sigma^2\neq\sigma_0^2.$$

现有

$$\prod_{i=1}^n f(x_i,\sigma)=\frac{1}{(\sqrt{2\pi})^n\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-a_0)^2\right\},$$

其中

$$T(x_1,\cdots,x_n)=\sum_{i=1}^n(x_i-a_0)^2;$$

$$Q(\sigma^2)=-\frac{1}{2\sigma^2};$$

$$h(x_1,\cdots,x_n)=\frac{1}{(\sqrt{2\pi})^n};$$

$$R(\sigma^2)=\frac{1}{(\sigma^2)^{n/2}}.$$

因此

$$\varphi(x_1,\cdots,x_n)=\begin{cases}1, & T<c_1 \quad 或 \ >c_2;\\ b_j, & T=c_j \quad j=1,2;\\ 0, & c_1<T<c_2.\end{cases} \tag{8.4.31}$$

由(8.4.30)式知

$$\alpha=M(\sigma_0^2,\sigma_0^2)=E_{\sigma_0^2}[\varphi(\xi_1,\cdots,\xi_n)];$$

$$\alpha E_{\sigma_0^2}\left[\sum_{i=1}^n(\xi_i-a_0)^2\right]=E_{\sigma_0^2}\left[\varphi(\xi_1,\cdots,\xi_n)\cdot\sum_{i=1}^n(\xi_i-a_0)^2\right], \tag{8.4.32}$$

我们知道，当 H_0 为真时

$$\frac{1}{\sigma_0}\sum_{i=1}^n(\xi_i-a_0)^2\sim\chi_n^2.$$

于是(8.4.32)式化简为

$$\alpha E_{\sigma_0^2}\left[\sum_{i=1}^n(\xi_i-a_0)^2\right]=\alpha\cdot n\sigma_0^2;$$

$$P_{\sigma_0^2}\left\{\sum_{i=1}^{n}(\xi_i-a_0)^2=c_j\right\}=0 \quad (j=1,2);$$

$$\alpha=E_{\sigma_0^2}[\varphi(\xi_1,\cdots,\xi_n)]=P_{\sigma_0^2}\{T<c_1\}+P_{\sigma_0^2}\{T>c_2\}.$$

记 $f_{\sigma_0^2}(y)$ 为统计量 $T=\sum_{i=1}^{n}(\xi_i-a_0)^2$ 在 H_0 为真时的密度函数为:

$$f_{\sigma_0^2}(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)(\sigma_0^2)^{n/2}}y^{\frac{n}{2}-1}\exp\left\{\dfrac{-y}{2\sigma_0^2}\right\}, & y\geqslant 0;\\ 0 & , y<0,\end{cases}$$

则(8.4.32)式又可化简为:

$$\int_{c_1}^{c_2}f_{\sigma_0^2}(y)\,dy=1-\alpha, \tag{8.4.33}$$

并有

$$\begin{aligned}\alpha\cdot n\sigma_0^2&=\int_{-\infty}^{c_1}yf_{\sigma_0^2}(y)\,dy+\int_{c_2}^{\infty}yf_{\sigma_0^2}(y)\,dy\\&=n\sigma_0^2-\int_{c_1}^{c_2}yf_{\sigma_0^2}(y)\,dy.\end{aligned}$$

即有

$$\int_{c_1}^{c_2}yf_{\sigma_0^2}(y)\,dy=(1-\alpha)n\sigma_0^2;$$

$$\begin{aligned}1-\alpha&=\int_{c_1}^{c_2}\frac{y}{n\sigma_0^2}f_{\sigma_0^2}(y)\,dy\\&=\int_{c_1}^{c_2}\frac{1}{2^{(n+2)/2}\Gamma\left(\frac{n+2}{2}\right)(\sigma_0^2)^{(n+2)/2}}y^{\frac{n+2}{2}-1}\exp\left\{\frac{-y}{2\sigma_0^2}\right\}dy\\&=\int_{c_1}^{c_2}f_{\sigma_0^2}(z)\,dz,\end{aligned}$$

其中 $f_{\sigma_0^2}(z)$ 为服从 $\chi^2_{(n+2)}$-分布的统计量 z 的密度函数.

$$\int_{c_1}^{c_2}f_{\sigma_0^2}(z)\,dz=\int_{c_1}^{c_2}f_{\sigma_0^2}(y)\,dy=1-\alpha,$$

应用分部积分法得:

$$\begin{aligned}\int_{c_1}^{c_2}f_{\sigma_0^2}(y)\,dy&=\{yf_{\sigma_0^2}(y)\}\Big|_{c_1}^{c_2}-\int_{c_1}^{c_2}yf'_{\sigma_0^2}(y)\,dy\\&=\frac{1}{2^{n/2}\Gamma(n/2)(\sigma_0^2)^{n/2}}y^{\frac{n}{2}}e^{-\frac{y}{2\sigma_0^2}}\Big|_{c_1}^{c_2}\end{aligned}$$

$$-\left(\frac{n}{2}-1\right)\int_{c_1}^{c_2} f_{\sigma_0^2}(y)dy+\frac{n}{2}\int_{c_1}^{c_2} f_{\sigma_0^2}(z)dz$$

$$=\frac{1}{2^{n/2}\Gamma(n/2)(\sigma_0^2)^{n/2}}y^{\frac{n}{2}}e^{-\frac{y}{2\sigma_0^2}}\Bigg|_{c_1}^{c_2}+\int_{c_1}^{c_2} f_{\sigma_0^2}(y)dy,$$

因此得

$$c_2^{\frac{n}{2}}e^{-\frac{c_2}{2\sigma_0^2}}=c_1^{\frac{n}{2}}e^{-\frac{c_1}{2\sigma_0^2}}. \tag{8.4.34}$$

给定检验大小 α 值，由(8.4.33)及(8.4.34)式，可确定 c_1 及 c_2 值，从而确定(8.4.31)式为

$$\varphi(x_1,\cdots,x_n)=\begin{cases}1, & T\leqslant c_1 \text{ 或} \geqslant c_2;\\ 0, & c_1<T<c_2.\end{cases}$$

例 8.4.7 设总体 ξ 服从具有参数为 $\lambda>0$ 的泊松分布，$\xi_1,\cdots,\xi_n$ 为其子样．考虑检验问题：

$$H_0:\quad \lambda=\lambda_0,\quad H_1:\quad \lambda\neq\lambda_0.$$

现有

$$\prod_{i=1}^{n}P(x_i,\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=e^{-n\lambda}\lambda^{\sum\limits_{i=1}^{n}x_i}\frac{1}{\prod\limits_{i=1}^{n}(x_i!)}$$

$$=e^{-n\lambda}\cdot\exp\left\{\left(\sum_{i=1}^{n}x_i\right)\ln\lambda\right\}\frac{1}{\prod\limits_{i=1}^{n}(x_i!)},$$

其中

$$T(x_1,\cdots,x_n)=\sum_{i=1}^{n}x_i;$$

$$Q(\lambda)=\ln\lambda;$$

$$h(x_1,\cdots,x_n)=1\Big/\prod_{i=1}^{n}(x_i!);$$

$$R(\lambda)=e^{-n\lambda}.$$

因此

$$\varphi(x_1,\cdots,x_n)=\begin{cases}1, & T<c_1 \text{ 或 } >c_2;\\ b_j, & T=c_j,\ j=1,2;\\ 0, & c_1<T<c_2.\end{cases} \tag{8.4.35}$$

由(8.4.30)式知

$$\left.\begin{aligned}&\alpha E_{\lambda}\left[\sum_{i=1}^{n}\xi_i\right]=E_{\lambda_0}\left[\varphi(\xi_1,\cdots,\xi_n)\cdot\sum_{i=1}^{n}\xi_i\right];\\&\alpha=M(\mathscr{K}_0,\lambda_0)=E_{\lambda_0}[\varphi(\xi_1,\cdots,\xi_n)].\end{aligned}\right\}\qquad(8.4.36)$$

我们知道，当 H_0 为真时，统计量 $T=\sum_{i=1}^{n}\xi_i$ 服从参数为 $n\lambda_0$ 的泊松分布，$E(T)=n\lambda_0$. 由(8.4.36)推得

$$\alpha=\sum_{j=0}^{c_1-1}\frac{(n\lambda_0)^j}{j!}e^{-n\lambda_0}+\sum_{j=c_2+1}^{\infty}\frac{(n\lambda_0)^j}{j!}e^{-n\lambda_0}$$

$$+\frac{b_1}{c_1!}(n\lambda_0)^{c_1}e^{-n\lambda_0}+\frac{b_2}{c_2!}(n\lambda_0)^{c_2}e^{-n\lambda_0};$$

$$n\lambda_0\alpha=\sum_{j=0}^{c_1-1}\frac{j(n\lambda_0)^j}{j!}e^{-n\lambda_0}+\sum_{j=c_2+1}^{\infty}\frac{j(n\lambda_0)^j}{j!}e^{-n\lambda_0}$$

$$+\frac{b_1c_1}{c_1!}(n\lambda_0)^{c_1}e^{-n\lambda_0}+\frac{b_2c_2}{c_2!}(n\lambda_0)^{c_2}e^{-n\lambda_0}.$$

在一般情形下上述方程不那么容易求解. 但当 λ_0 不接近于零而 n 又不很小时，统计量

$$U=\frac{\sum_{i=1}^{n}\xi_i-n\lambda_0}{\sqrt{n\lambda_0}}$$

有渐近分布，它为正态 $N(0,1)$，那么

$$P_{\lambda_0}\left\{\sum_{i=1}^{n}\xi_i-n\lambda_0<-u\right\}=P_{\lambda_0}\left\{\sum_{i=1}^{n}\xi_i-n\lambda_0>u\right\}.$$

对一切实数 u 都渐近地成立(这是因为正态分布具有对称性). 因此，b_1、b_2 及 c_1、c_2 由下式确定:

$$\frac{\alpha}{2}=\sum_{j=0}^{c_1-1}\frac{(n\lambda_0)^j}{j!}e^{-n\lambda_0}+\frac{b_1}{c_1!}(n\lambda_0)^{c_1}e^{-n\lambda_0}$$

$$=\sum_{j=c_2+1}^{\infty}\frac{(n\lambda_0)^j}{j!}e^{-n\lambda_0}+\frac{b_2}{c_2!}(n\lambda_0)^{c_2}e^{-n\lambda_0}.$$

特别地，如果 $n=10$，$\lambda_0=1$，$\alpha=0.05$，则有

$$\sum_{j=0}^{8}\frac{(10)^j e^{-10}}{j!}=0.010,$$

$$\sum_{j=0}^{4}\frac{(10)^j e^{-10}}{j!}=0.029,$$

$$\sum_{j=0}^{17}\frac{(10)^j e^{-10}}{j!}=0.986,$$

$$\sum_{j=0}^{16}\frac{(10)^j e^{-10}}{j!}=0.973.$$

于是得到

$$c_1=4,\ \ c_2=16;\ \ b_1=0.789,\ \ b_2=0.846.$$

*§8.5 质量控制

在工业生产过程中常常采用一种叫“质量控制”的统计方法，以检查和控制产品的质量. 特别是现代工业的自动化高速生产，怎样对整个生产过程进行控制，显得很重要. 质量控制这种统计方法，通常具有下列两方面的工作：

(1) 工序控制——其目的是在生产过程中通过检查产品质量以及时发现问题，如预告生产稳定性即将破坏、废品即将出现，事故即将发生等. 并能在一定程度上指出毛病所在，以便及时采取措施. 这种质量控制是积极性的，称它为预告性的质量检查.

(2) 验收控制——在生产出一批产品后，制定检查方案对产品质量加以鉴定，合乎规格的可以出厂或验收；不合规格的作废品，并采取措施改进生产. 这是鉴定性的质量检查.

验收控制的主要内容实际上是讨论“抽样检查方案”怎样制订并较为合理，因篇幅所限，请参看[5]的第 14 章. 一种更完善的检验方法是 A. Wald 所建立的序贯分析. 关于这方面的详细叙述可

参阅他的专著. 也可参看[5]的第17章.

现在，只对工序控制作扼要的讨论. 我们说生产过程是稳定的，是指如果影响产品质量的主要因素如原料、工艺、机器、人力等无显著变化时，那么产品质量也不会发生很大的差异. 虽然如此，即使生产过程是稳定的，而在生产过程中，还会有许多随机因素在起作用. 因此，我们说生产是稳定的，或者说生产过程是处于控制状态中，是指生产的基本条件不变而排斥不了随机性因素在起作用，这时产品质量的变化遵循着某种概率统计规律性. 反之，生产出现不稳定，是指出现了非随机因素在起作用，或者生产的规律性受到破坏，出现事故. 工序控制的目的，就是要检查生产过程是不是处于控制状态中，如果出现生产违反规律性的状态时，应立即采取技术措施以保证产品的质量.

要进行工序控制，首先要了解产品质量的指标 ξ 服从什么分布. 如钢条的直径 ξ、汽车的汽缸螺栓的口径 ξ，一般都服从正态分布；一批电灯泡中的废品个数 ξ，一批滚珠中的废品个数 ξ、一般都服从二项分布；又如铸件表面上的气孔或砂眼的个数 ξ，棉布面上的疵点个数 ξ，一般都服从泊松分布.

例 8.5.1 某汽车零件制造厂的车床加工汽缸螺栓，试对"螺栓口径"制订质量控制图.

解 记"螺栓口径"为 ξ，假定总体 ξ 服从正态 $N(a,\sigma)$，所谓"生产过程处于稳定状态"，是指总体 ξ 的两个参数 a 及 σ^2 在规定的标准尺寸及允许的"公差"范围之内. 因此，当总体 ξ 是服从正态时，质量控制问题就归结为控制一对参数 a 及 σ^2 在规定的标准尺寸及允许的"公差"范围之内.

下面结合例 8.5.1 介绍平均值控制图及极差控制图（极差可作为 σ 的估计，所以用极差控制图方便）.

一、平均值控制图

1. 基本思想

当总体 ξ 服从正态 $N(a,\sigma)$ 的时候，则子样平均 $\bar{\xi}$ 服从正态 $N\left(a,\frac{\sigma}{\sqrt{n}}\right)$，若 σ 由以往经验能确定并记作 σ_0，则给定置信水平 $1-\alpha$（α 由具体要求定）. 由(7.4.2)知有

$$P\left\{a-u_\alpha\frac{\sigma_0}{\sqrt{n}}\leqslant\bar{\xi}\leqslant a+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right\}=1-\alpha. \qquad (8.5.1)$$

对于子样 $\xi_1,\cdots,\xi_n$ 的观察值 $x_1,\cdots,x_n$，计算得平均值

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,$$

若 $\bar{x}\in\left(a-u_\alpha\frac{\sigma_0}{\sqrt{n}},a+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right)$，则以置信水平 $1-\alpha$ "认为"生产过程处于稳定状态，否则就要采取维修措施，以保证生产正常. 其中 a 为规定的标准尺寸，把

$$\left(a-u_\alpha\frac{\sigma_0}{\sqrt{n}},a+u_\alpha\frac{\sigma_0}{\sqrt{n}}\right)$$

称为子样平均 $\bar{\xi}$ 的置信水平为 $1-\alpha$ 的控制域，它是由两条平行线组成的.

如果 σ^2 未知而要用 S^2 作估计，则由(7.4.3)知有

$$P\left\{a-t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\leqslant\bar{\xi}\leqslant a+t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\right\}=1-\alpha. \qquad (8.5.2)$$

这时 $\bar{\xi}$ 的置信水平为 $1-\alpha$ 的控制域是 $\left(a-t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}},\right.$ $\left.a+t_{(n-1)}(\alpha)\frac{S}{\sqrt{n-1}}\right)$，这是由两条斜线组成的区域.

2. 基本作法

由(8.5.1)看到，对于不同的置信水平，有不同的控制区域. 例如$\alpha_1=0.003$，有$u_{\alpha_1}=3$，则控制区域为$\left(a-\frac{3\sigma_0}{\sqrt{n}}, a+\frac{3\sigma_0}{\sqrt{n}}\right)$. 以直线$y=a$为中心，作两条平行线：

$$L_1:\ \ y=a+3\frac{\sigma}{\sqrt{n}};$$

$$L_2:\ \ y=a-3\frac{\sigma}{\sqrt{n}},$$

则称以直线L_1与L_2所作的图为子样平均$\bar{\xi}$的控制图. 又如给定$\alpha_2=0.0045$，那么$u_{\alpha_2}=2$，又可作两条平行线：

$$M_1:\ \ y=a+2\frac{\sigma}{\sqrt{n}};$$

$$M_2:\ \ y=a-2\frac{\sigma}{\sqrt{n}},$$

如图(8.5.1)所示.

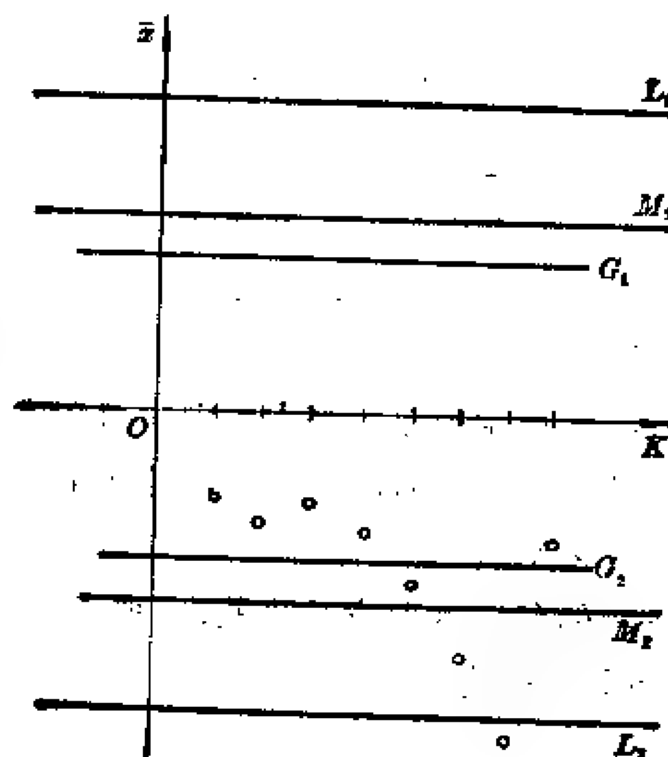


图 8.5.1

假定我们规定：当子样平均ξ的观察值$\bar{x}$落在L_1与L_2所围成的区域之外时，出现废品，认为生产不稳定．当$\bar{x}$落在M_1与M_2所围成的区域之外时，表明生产稳定有破坏的危险，即M_1与M_2起着警告作用；$\bar{x}$值落在M_1与M_2所围成的区域之内，生产保持稳定，认为生产可继续进行或认为产品合格．

在这种质量控制图中，称M_1、M_2分别为上、下控制线，L_1与L_2分别为上、下废品线．

现对一台车床加工的汽缸螺栓测其口径，每天抽查八次；每隔一小时抽一次，每次抽15只($n=15$)，测其口径，(至百分之一毫米)，算出其平均值$\bar{x}$．每天有八个平均值$\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_8$，依次点在控制图上，如上图8.5.1所示，横坐标为子样的标号，纵坐标为ξ的观察值$\bar{x}$．第六次抽得的$\bar{x}$值落在M_2与L_2之间，发出警告；第七次抽得的$\bar{x}$值落在L_2之下，出现废品．就是说，发出警告后不采取措施，废品即将出现．在出现废品后，采取了维修措施，从第八次抽得的平均值$\bar{x}$落在M_1与M_2区域内，表明生产稳定已恢复．

利用"平均值质量控制图"不必等到点落在控制区域之外，往往就能看出倾向性的问题．例如，第一，如果点总是落在中心线$y=a$以上(或以下)，或者点有单调上升(或下降)，这可能是车床或零件安装不适当(太紧或太松)，或者是机件、工具已受损耗、磨损；第二，点在控制区域内波动很大，一会儿靠近上控制线，一会儿又靠近下控制线，这可能是原料等主要因素有较大的波动；第三，如果点的联线有某种规律(如周期现象)，这也能启发我们及时注意车床可能出现什么毛病．

如果生产过程的稳定性要求严格些，可取$\alpha_0=0.10$，那么$u_\alpha=1.65$，相应地作两条平行线：

$$G_1: \ y=a+1.65\frac{\sigma}{\sqrt{n}};$$

$$G_2: \quad y = a - 1.65\frac{\sigma}{\sqrt{n}},$$

G_1 及 G_2 都在 M_1 与 M_2 所围成的区域之内，我们以 M_1、与 M_2 分别为上、下废品线，以 G_1, G_2 分别为上、下控制线. 此时如上图(8.5.1)所示，第五次抽得的点落在 G_2 的下面，发出警告，而第六次抽得的点表明废品出现了.

当然，在生产过程为稳定的情况下，子样平均值 $\bar{\xi}$ 的观察值 $\bar{x}$ 也有可能落在控制线围成的区域之外，出现"虚报警告"的现象，浪费一些时间和物力作不必要的维修，这种"虚报警告"事件出现的概率为 α. 如使用 M_1——M_2 控制线，"虚报警告"事件出现的概率为 $\alpha=0.045$，而使用 G_1——G_2 控制线，"虚报警告"事件出现的概率为 $\alpha=0.10$. 可见，生产的稳定性要求越高，控制区域越窄，而"虚报警告"的可能性就越大. 怎样"权衡得失"，这种思想方法已在前面的"最佳检验法"(§8.4)中作了阐述.

二、极差控制图

由(7.4.4)式知，有

$$P\left\{\sigma^2\chi^2_{(n-1)}\left(1-\frac{\alpha}{2}\right)\leqslant nS^2\leqslant\sigma^2\chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)\right\}$$
$$=1-\alpha, \tag{8.5.3}$$

给定允许"公差"范围，可定 σ^2 的数值 σ_0^2，从而对子样方差 S^2 作出控制图. 在应用(8.5.2)及(8.5.3)两式作控制图时，要求每次计算子样方差 S^2 的观察值，而它远比计算子样极差 D_n^* 的观察值繁琐得多. 由于找到了子样极差 D_n^* 同总体方差 σ^2 的关系式(6.2.23)，并提供了表(6.2.1)，因而极差控制图很受欢迎. 设总体 ξ 服从正态 $N(a,\sigma)$，$\xi_1,\cdots,\xi_n$ 为其子样，$D_n^*=\max\limits_{1\leqslant i\leqslant n}\{\xi_i\}-\min\limits_{1\leqslant i\leqslant n}\{\xi_i\}$ 为子样极差，则有：

$$E(D_n^*)=c_n\sigma, D(D_n^*)=v_n^2\sigma^2.$$

近似地应用(8.5.1)式,有

$$P\{E(D_n^*)-u_\alpha\sqrt{D(D_n^*)}\leqslant D_n^*\leqslant E(D_n^*)+u_\alpha\sqrt{D(D_n^*)}\}$$
$$=1-\alpha,$$

如给定 $\alpha=0.003$,查得 $u_\alpha=3$,从而有:

$$P\{c_n\sigma-3v_n\sigma\leqslant D_n^*\leqslant c_n\sigma+3v_n\sigma\}=0.997,$$

即

$$P\{(c_n-3v_n)\sigma\leqslant D_n^*\leqslant(c_n+3v_n)\sigma\}=0.997.$$

将给定的允许"公差"σ_0 作为 σ 值，对于容量为 n 的子样，在表(6.2.1)中查得 c_n 及 v_n 各值,这样就制定出容量为 n 的子样的极差 D_n^* 的控制图,上限及下限分别为:

$$R_1:\ y=(c_n+3v_n)\sigma_0;$$
$$R_2:\ y=(c_n-3v_n)\sigma_0.$$

实际工作是这样进行的:由容量为 n 的子样观察值 $x_1,\cdots,x_n$,计算得子样极差 D_n^* 的观察值 d_n,若 d_n 落在 R_1 与 R_2 围成的控制区域内,则认为生产过程处于稳定状态,否则就发出警告. 注意极差总是非负的,则 $y\leqslant 0$ 的下控制线可以取消. 但是，在一般情况下不能取消下控制线,子样极差 D_n^* 的观察值 d_n 并非越小越好,因为有时测量仪器不灵敏导致 d_n 很小,有时则是生产使用的原料质量过高所致. 这时,也应注意检查生产情况,对于后一种情形，工厂从经济核算方面来看未必合算.

*§8.6 子样容量 n 的确定

子样 $\xi_1,\cdots,\xi_n$ 的容量 n 的确定，在许多科学实验工作中是经常遇到的问题. 我们知道,容量 n 太小,估计问题不那么精确，检验问题不那么可靠；容量 n 太大，又会造成人力物力的浪费. 在此,我们所指的子样都是简单子样. 下面,我们就几种场合遇到的子样容量 n 的问题,作扼要的讨论.

一、参数估计与检验中 n 的确定

假定总体 ξ 服从正态 $N(a,\sigma)$，其中 a 及 σ 都是未知参数，给定显著性水平 α，要判断原假设 H_0: $a=a_0$. 当 H_0 成立时，由(8.2.2)知，

$$\sqrt{n-1}\frac{\bar{\xi}-a_0}{S}\sim t_{(n-1)}.$$

选择临界值 $t_{(n-1)}(\alpha)$，使得

$$P\left\{\left|\sqrt{n-1}\frac{\bar{\xi}-a_0}{S}\right|\geqslant t_{(n-1)}(\alpha)\right\}=\alpha, \tag{8.6.1}$$

亦即

$$P\left\{\left|\sqrt{n-1}\frac{\bar{\xi}-a_0}{S}\right|\leqslant t_{(n-1)}(\alpha)\right\}=1-\alpha,$$

其中

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i,\quad S^2=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2.$$

(8.6.1)式可以回答三方面的问题：一是当 H_0 成立时，给出了 H_0 的否定域，二是在 a 为已知时，给出了子样平均 $\bar{\xi}$ 的控制图，三是在 a 为未知时，给出了 a 的置信水平为 $1-\alpha$ 的区间估计

$$\left(\bar{\xi}-\frac{S}{\sqrt{n-1}}t_{(n-1)}(\alpha),\ \bar{\xi}+\frac{S}{\sqrt{n-1}}t_{(n-1)}(\alpha)\right).$$

这个置信区间的长度记作 2Δ，其中

$$\Delta=\frac{S}{\sqrt{n-1}}t_{(n-1)}(\alpha). \tag{8.6.2}$$

在实际工作中，常称 Δ 为估计精度或误差精度、试验精度.

在(8.6.2)式中，如果事先给定 Δ 值，则有

$$n=1+\frac{S^2}{\Delta^2}t_{(n-1)}^2(\alpha). \tag{8.6.3}$$

又当总体的方差 σ^2 由以往经验能确定为 σ_0^2，用 σ_0^2 代替 S^2；给定

显著性水平 α，能查得临界值 $t_{(n-1)}(\alpha)$. 这时，万事俱备，n 值才能由(8.6.3)式所确定. 但是，确定临界值 $t_{(n-1)}(\alpha)$ 本身，事先就要知道 n 值，即自由度 $n-1$ 的值. 因此，(8.6.3)式还没有真正解决计算 n 值的问题. 然而，我们从 t 分布临界值表中看到，对于显著性水平 $\alpha\leqslant 0.05$ 的情形，当 $n>30$ 时，其临界值 $t_{(n-1)}(\alpha)\approx 2$，这个临界值对于大于 30 的各个 n 值的影响不太大. 因此，我们先采用近似公式

$$n=1+\frac{4S^2}{\Delta^2} \tag{8.6.4}$$

计算 n. 当计算出的 n 大大超过 30 时，这就与 $t_{(n-2)}(\alpha)\approx 2$ 不矛盾了.

在实际工作中是这样确定 n 的：给定 S 及 Δ 的值，按(8.6.4)式计算 n，若 n 值大于 30，就以这个 n 值作为子样的容量；若 n 值不大于 30，则采用一种叫“试差法”来确定子样的容量 n. 试差法是这样：先由(8.6.4)式计算出的 n，作为第二次查临界值 $t_{(n-1)}(\alpha)$ 时用的 n，将查得的临界值 $t_{(n-1)}(\alpha)$ 代入(8.6.3)式计算 n 值；再由这个 n 值作为第三次查临界值 $t_{(n-1)}(\alpha)$ 时用的 n，又将查得的临界值 $t_{(n-1)}(\alpha)$ 再代入(8.6.3)式计算 n 值；如此循环，直至(8.6.3)式中两边的 n 值相同或差异很小为止. 一般试数次就行了，计算出的 n 值通常不要小于 5.

例 8.6.1 某种电子仪器额定电流的总体的方差 $\sigma^2\approx 200$，经检验确定的置信水平 0.95 的置信区间的长度为 20mA，问要检查多少台仪器，才能满足这样的要求？

解 取 $S^2=200$，$\Delta=10$，则由(8.6.4)式计算：

$$n=1+\frac{800}{100}=9.$$

由于 n 值小于 30，所以要用“试差法”. 对于 $\alpha=0.05$，查得临界值

$t_8(0.05)=2.31$，代入(8.6.3)式算得

$$n=1+\frac{200}{100}\times(2.31)^2=11.67.$$

以自由度 11 及 $\alpha=0.05$，查得临界值 $t_{11}(0.05)=2.20$，再代入(8.6.3)式，计算得

$$n=1+\frac{200}{100}\times(2.20)^2=10.68.$$

以自由度 10 及 $\alpha=0.05$，查得临界值 $t_{10}(0.05)=2.23$，再代入(8.6.3)式，计算得 $n=10.94$. 这时确定 $n=11$，不必再"试差"了.

在这个例子中，如果事先给定 $\Delta=5\text{mA}$，α 及 S^2 不变，则由(8.6.4)式计算得

$$n=1+\frac{4\times200}{25}=33.$$

这时不必采用"试差法"了，就确定子样容量为 $n=33$. 其直观意义是：置信区间的长度缩短了，置信水平 $1-\alpha$ 不减小，唯有子样容量 n 增大才行.

例 8.6.2 为了判别某种药剂对高血压的疗效，治疗前与治疗后，血压差数的方差一般为 15左右，即 $\sigma^2\approx15$. 假定该药有效，要求在显著性水平 $\alpha=0.05$ 下能辨出误差精度 $\Delta=5\text{mHg}$ 的血压差，问需多少个病例的试验？

解 本例作为两个总体的数学期望是否相等的配对比较试验，同例 8.2.5 之(1)那样，变为单个总体情形的方法来确定 n. 用 d 表示治疗前后的血压差，取 $S_d^2=15$，$\Delta_d=5$，$\alpha=0.05$，由(8.6.4)式计算得

$$n=1+\frac{4S_d^2}{\Delta_d^2}=3.4.$$

因 n 值太小，则取 $n=5$ 起，采用"试差法"确定 n：由 $\alpha=0.05$，$n=5$，查得临界值 $t_4(0.05)=2.78$，代入(8.6.3)式计算得 $n=5.6$. 今

取 $n=6$，对于 $\alpha=0.05$，查得临界值 $t_5(0.05)=2.31$，再代入(8.6.3)式计算得 $n=4$. 由于 n 值小于 5，则定 $n=5$，即做 5 个病例的试验就行了.

现在对于两个总体在不配对比较试验中 n 的确定问题讨论如下：

假定总体 ξ 及 η 分别服从正态 $N(a_1,\sigma)$ 及 $N(a_2,\sigma)$，其中 a_1，a_2 及 σ 都是未知参数，在显著性水平 α 下，要判断假设：

$$H_0:\ a_1=a_2,\quad H_1:\ a_1\neq a_2.$$

由(8.2.6)式知，(a_1-a_2) 的置信水平为 $1-\alpha$ 的置信区间为：

$$\left(\bar{\xi}-\bar{\eta}-S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{(n_1+n_2-2)}(\alpha),\right.$$

$$\left.\bar{\xi}-\bar{\eta}+S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{(n_1+n_2-2)}(\alpha)\right),$$

其中 $S^2=\dfrac{1}{n_1+n_2-2}(n_1S_1^2+n_2S_2^2)$. 置信区间的长度记作 2Δ，则 Δ 为

$$\Delta=S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{(n_1+n_2-2)}(\alpha). \tag{8.6.5}$$

我们仅就 $n_1=n_2=n$ 的情形作讨论，这时试验精度 Δ 为

$$\Delta=S\sqrt{\frac{2}{n}}t_{2(n-1)}(\alpha), \tag{8.6.6}$$

从而有

$$n=\frac{2S^2}{\Delta^2}t_{2(n-1)}^2(\alpha). \tag{8.6.7}$$

对于 $\alpha=0.05$，当 $n>15$ 时，$t_{2(n-1)}(\alpha)\approx 2$，于是

$$n\approx\frac{8S^2}{\Delta^2}. \tag{8.6.8}$$

由(8.6.4)同(8.6.8)作比较看到，对于两个总体的数学期望是否相等的检验问题，在相同的试验精度下，用配对比较试验，将比不

配对比较试验所需的子样容量大约减少一半，因而在试验条件许可的情况下，尽量采用配对比较试验为好.

例 8.6.3 若例 8.6.2 中的 Δ、σ^2 值不变，今用不配对比较试验，试问子样容量 n 应多大？

解 先由(8.6.8)式计算得

$$n=\frac{8\times 15}{5^2}=4.8.$$

取定 $n=5$. 对于显著性水平 $\alpha=0.05$, $n=5$, 查得临界值 $t_8(0.05)=2.31$, 代入(8.6.7)式，计算得

$$n=\frac{2\times 15}{5^2}(2.31)^2=6.4.$$

取定 $n=6$, 对于显著性水平 $\alpha=0.05$, 查得临界值 $t_{10}(0.05)=2.23$ 再代入(8.6.7)式，计算得

$$n=\frac{2\times 15}{5^2}(2.23)^2=6,$$

最后确定 $n=6$. 就是说，不用配对比较试验，那么两个子样是各为 6 个病例.

二、最佳检验中 n 的确定

假设总体 ξ 服从正态 $N(a,\ \sigma)$, 其中 a 为未知参数，σ 为已知，记作 σ_0, 提出假设检验问题：

$$H_0:\ a=a_0,\quad H_1:\ a=a_1,$$

其中 $a_0<a_1$. 在给定犯两类错误的概率 α 及 β 的大小，用最佳检验法判断这个假设时，试问子样容量 n 应多大？

我们知道，对于上述假设检验问题，最佳否定域为

$$\bar{\xi}\geqslant A. \tag{8.6.9}$$

当原假设 H_0 为真时，子样平均 $\bar{\xi}$ 服从正态 $N\left(a_0,\ \frac{\sigma_0}{\sqrt{n}}\right)$; 若 H_0 不为真，即备选假设 H_1 为真，则子样平均 $\bar{\xi}$ 服从正态 $N\left(a_1,\ \frac{\sigma_0}{\sqrt{n}}\right)$.

因而有方程组：

$$\begin{cases}\displaystyle\int_A^{\infty}\frac{1}{\sqrt{2\pi}\,(\sigma_0/\sqrt{n})}e^{-\frac{(\bar{x}-a_0)^2}{2(\sigma_0/\sqrt{n})^2}}d\bar{x}=\alpha;\\ \displaystyle\int_{-\infty}^{A}\frac{1}{\sqrt{2\pi}\,(\sigma_0/\sqrt{n})}e^{-\frac{(\bar{x}-a_1)^2}{2(\sigma_0/\sqrt{n})^2}}d\bar{x}=\beta.\end{cases}\tag{8.6.10}$$

将子样平均$\bar{\xi}$标准正态化，作变换

$$U=\frac{\bar{\xi}-a}{\sigma_0/\sqrt{n}},$$

U 服从正态 $N(0,1)$，则(8.6.10)方程组等价于

$$\begin{cases}\displaystyle\int_{\frac{A-a_0}{\sigma_0/\sqrt{n}}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=\alpha;\\ \displaystyle\int_{-\infty}^{\frac{A-a_1}{\sigma_0/\sqrt{n}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du=\beta.\end{cases}\tag{8.6.11}$$

在标准正态分布表中，给定α，查得数值 $u_\alpha=\dfrac{A-a_0}{\sigma_0/\sqrt{n}}$，使得

$$p\{U\geqslant u_\alpha\}=\alpha.$$

给定β，查出 $u_\beta=\dfrac{A-a_1}{\sigma_0/\sqrt{n}}$，使得

$$p\{U\leqslant u_\beta\}=\beta$$

于是

$$u_\alpha\cdot\frac{\sigma_0}{\sqrt{n}}+a_0=u_\beta\cdot\frac{\sigma_0}{\sqrt{n}}+a_1,\tag{8.6.12}$$

从而解得

$$\left.\begin{aligned}n&=\frac{(u_\alpha-u_\beta)^2\cdot\sigma_0^2}{(a_1-a_0)^2}\\ A&=\frac{a_1u_\alpha-a_0u_\beta}{u_\alpha-u_\beta}\end{aligned}\right\}\tag{8.6.13}$$

举例说明：若 $a_0=0, a_1=1, \sigma_0=1, \alpha=\beta=0.05$，则在标准正态分布表中查得

$$u_\alpha = 1.645, \quad u_\beta = -1.645.$$

由(8.6.13)式计算得

$$n = 10.8, \quad A = 0.5.$$

就是说，子样容量 $n = 11$，最佳否定域为 $\bar{\xi} \geqslant 0.5$.

三、验收抽样方案中 n 的确定

例 **8.6.4** 今要验收一批水泥：如果这种水泥制成混凝土后，断裂强度为 5000(单位)，验收者希望 100 次中有 95 次被"接收"；若断裂强度为 4600(单位)，验收者希望 100 次中只通过 10 次. 已知混凝土断裂强度的均方差 $\sigma = 600$(单位)，试为验收者制定验收抽样方案.

解 这种验收抽样方案，在抽样理论中，称之为单式计量验收方案. 设计这种验收方案时，是以总体 ξ 的数学期望 $E(\xi) = a$ 作为验收的质量指标. 现在，我们在给定犯两种类型错误的概率 α 及 β 下，讨论验收抽样方案的设计：定出子样容量 n 的大小，确定对一批成品"接收"或"拒收"的标准 L 及 H(在序列抽样理论中，凭借子样作出验收结果时，是分三种情况"接收"、"拒收"及"待定"，请参看[5]的第 17 章).

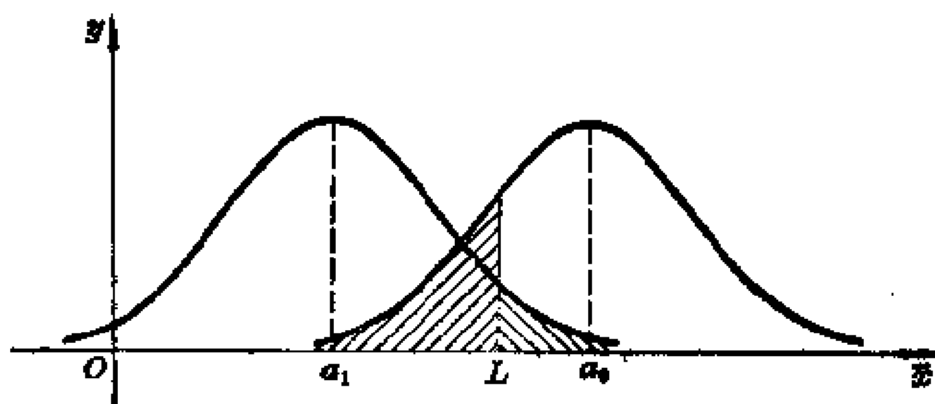


图 8.6.1

假定总体 ξ 服从正态 $N(a, \sigma)$，那么子样平均 $\bar{\xi}$ 就服从正态

$N\left(a, \frac{\sigma}{\sqrt{n}}\right)$，在本例中 $a_0=5000, a_1=4600, \sigma_0=600$．就是说断裂强度达到 a_0 时，成品合格，低于 a_1 时，成品不合格．α 是成品合格时而作出拒收错误的概率，β 是成品不合格时而作出"接收"错误的概率，在本例中是取 $\alpha=0.05, \beta=0.10$．于是有方程组

$$\begin{cases} \int_{-\infty}^{L} \frac{1}{\sqrt{2\pi}\ (\sigma_0/\sqrt{n})} e^{-\frac{(\bar{x}-a_0)^2}{2(\sigma_0/\sqrt{n})^2}} d\bar{x} = 0.05; & (8.6.14) \\ \int_{L}^{\infty} \frac{1}{\sqrt{2\pi}\ (\sigma_0/\sqrt{n})} e^{-\frac{(\bar{x}-a_1)^2}{2(\sigma_0/\sqrt{n})^2}} d\bar{x} = 0.10. & (8.6.15) \end{cases}$$

将子样平均 ξ 标准正态化，则(8.6.14)式等价于

$$\int_{-\infty}^{\frac{L-a_0}{\sigma_0/\sqrt{n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 0.05;$$

$$\frac{L-a_0}{\sigma_0/\sqrt{n}} = -1.645, \tag{8.6.16}$$

(8.6.15)式等价于

$$\int_{\frac{L-a_1}{\sigma_0/\sqrt{n}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 0.10;$$

$$\frac{L-a_1}{\sigma_0/\sqrt{n}} = 1.282. \tag{8.6.17}$$

于是

$$L = a_0 - \frac{\sigma_0}{\sqrt{n}} \times 1.645 = a_1 + \frac{\sigma_0}{\sqrt{n}} \times 1.282,$$

从而解得

$$n = 19, \quad L = 4775. \tag{8.6.18}$$

就是说，在每批待验的成品中只需随机抽取 19 件进行检查．若混凝土的断裂强度达到或超过 4775(单位)，便"接收"这批成品，否则就不接收这批成品．

习 题

1. 某化工原料在处理前后取样分析，测得其含脂率的数据如下：

处理前	0.19	0.18	0.21	0.30	0.66
	0.42	0.08	0.12	0.30	0.27
处理后	0.19	0.24	1.04	0.08	0.20
	0.12	0.31	0.29	0.13	0.07

假定处理前后的含脂率都服从正态分布，且其标准差 σ 不变，给定显著性水平 $\alpha=0.05$，问处理前后含脂率的平均值有无显著性的变化.

2. 今有两台机床加工同一零件，分别取 6 个及 9 个零件测其口径，数据记为 $(x_1, \cdots, x_6)$ 及 $(y_1, \cdots, y_9)$，计算得：

$$\sum_{i=1}^{6} x_i = 204.6, \qquad \sum_{i=1}^{6} x_i^2 = 6978.93,$$

$$\sum_{j=1}^{9} y_j = 370.8, \qquad \sum_{j=1}^{9} y_j^2 = 15280.173.$$

假定零件口径 ξ 服从正态分布，给定显著性水平 $\alpha=0.05$，问是否可认为这两台机床加工零件口径的方差无显著性的差异？

3. 某电话站在一小时内接到电话用户的呼唤次数按每分钟记录得如下表：

呼唤次数	0	1	2	3	4	5	6	≥7
频数	8	16	17	10	6	2	1	0

试问这个分布能否看作为泊松分布？

4. 在某公路上 50 分钟之内，记录每 15 秒钟过路汽车的辆数，得到分布情况如下：

辆数	0	1	2	3	4	5
频数	92	68	28	11	1	0

试问这个分布能否看作为泊松分布.

5. 在数 $\pi=3.14159\cdots$ 的前 800 位小数中，数字 0、1、2、…、9 出现的次数记录如下：

数字	0	1	2	3	4	5	6	7	8	9
频数	74	92	83	79	80	73	77	75	76	91

试问这个分布能否看作为均匀分布.

6. 检查产品质量时，每次抽取 10 个产品来检查，共抽取 100 次，记录每 10 个产品中的次品数列如下表：

次品数	0	1	2	3	4	5	6	…	10
频数	35	40	18	5	1	1	0	…	0

试问生产过程中出现次品的概率能否看作是不变的，即次品数 ξ 是否服从二项分布？

7. 某香烟厂生产两种香烟，独立地随机抽取容量大小相同的烟叶标本测其尼古丁含量的毫克数，实验室分别作了六次测定，数据记录如下：

甲	25	28	23	26	29	22
乙	28	23	30	25	21	27

试问这两种香烟的尼古丁含量有无显著性差异？给定显著性水平 $\alpha=0.05$，假定含量服从正态分布并具有公共方差.

8. 为了研究慢性气管炎与吸烟量的关系，调查了三百八十五人，统计数字由下表所示.

类型 \ 人数 \ 烟量	a 支/日	b 支/日	c 支/日	求和
患病者人数	26	147	37	210
健康者人数	30	123	22	175
求和	56	270	59	385

试问慢性气管炎与吸烟量是否有关？给定显著性水平 $\alpha=0.05$.

9. 设总体 ξ 服从正态 $N(a,\sigma)$，考察如下检验问题：

$$H_0: a=a_0, \qquad H_1: a=a_1\ (\neq a_0).$$

证明当子样容量 n 充分大时，可使犯两类错误的概率任意地小.

10. 设总体ξ服从正态$N(a,2)$，$\xi_1,\cdots,\xi_{16}$为其子样，子样平均

$$\bar{\xi}=\frac{1}{16}\sum_{i=1}^{16}\xi_i,$$

考虑如下检验问题：

$$H_0: a=0,\quad H_1: a\neq 0.$$

试证下述三个否定域有相同的显著性水平$\alpha=0.05$：

(1) $2\bar{\xi}\leqslant -1.645$;

(2) $1.50\leqslant 2\bar{\xi}\leqslant 2.125$;

(3) $2\bar{\xi}\leqslant -1.96$ 及 $2\bar{\xi}\geqslant 1.96$.

11. 设总体ξ的密度函数为

$$f(x;\theta)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & 0<x<\infty,\ 0<\theta<\infty,\\ 0 & ,\text{其它}\end{cases}$$

$\xi_1,\cdots,\xi_n$为其子样，分别试求

(1) $H_0:\theta=2$, $H_1:\theta=4$;

(2) $H_0:\theta=2$, $H_1:\theta=1$

的最佳检验，给定显著性水平$\alpha=0.05$.

12. 设总体ξ服从贝努里分布$b(1,p)$，即有：

$$P(x;\ p)=\begin{cases}C_1^x p^x(1-p)^{1-x}, & x=0,1;\\ 0 & ,\text{其它},\end{cases}$$

$0<p<1$，$\xi_1,\cdots,\xi_n$为其子样，对于足够大的n，分别试求

(1) $H_0: p=\frac{1}{2}$, $\qquad H_1: p=\frac{1}{3}$;

(2) $H_0: p=\frac{1}{3}$, $\qquad H_1: p=\frac{1}{2}$

在显著性水平$\alpha=0.05$下的最佳检验.

13. 设总体ξ服从参数为λ的泊松分布：

$$P(x;\ \lambda)=\frac{\lambda^x}{x!}e^{-\lambda},\ x=0,1,2,\cdots,$$

$\lambda>0$，$\xi_1,\cdots,\xi_{10}$是容量为10的子样，试求

$$H_0:\lambda=0.1,\qquad H_1:\lambda=1$$

在显著性水平$\alpha=0.05$下的最佳检验.

14. 利用车贝谢夫不等式，试说明均匀对称的钱币需抛掷多少次，才能使得子样平均值 $\bar{\xi}$ 落在0.4到0.6之间的概率至少为0.9？

15. 对方差 σ^2 为已知的正态总体来说，问需抽取容量 n 为多大的子样，才能使得总体的数学期望 a 的置信水平 $1-\alpha$ 的置信区间的长度不大于 L？

16. 证明：若统计量 F 服从 F_{n_1,n_2} 分布，则 $\frac{1}{F}$ 服从 F_{n_2,n_1} 分布，且临界值之间有如下关系式：

$$F_{n_2,n_1}(1-\alpha)=\frac{1}{F_{n_1,n_2}(\alpha)}$$

第九章 回归分析与方差分析

我们在第七章的§7.7中所讲的最小均方误差估计原理，是在最小均方误差意义下，处理由一个随机变数来估计或预测另一个与之有相关关系的随机变数的问题.更一般地，是由一组随机变数来估计或预测另一个与它们具有一定的相关关系的随机变数. 研究这类问题的统计方法，称为多元相关分析. 在这一章中，我们将研究这类问题的一种情形，即一个随机变数同一个或一组非随机变数的关系问题. 这类问题的研究统称作回归分析. 先谈一个非随机变数 x 同一个随机变数 y 的关系问题，关于 x 的每一个值而对 y 作重复观察时，可能有几个不同的观察值，因而实际上是要在 y 的期望值 $E[y|x]$ 与 x 之间建立数学表达式. 若 $E[y|x]$ 是 x 的线性函数，如 $E[y|x]=\beta_0+\beta_1x$，这时所作的统计分析称为线性回归分析. 总之，由一个或一组非随机变数来估计或预测另一个随机变数的期望值时建立的线性模型及所作的统计分析，我们称之为线性回归分析. 处理这类问题的方法，通常是先假设这个随机变数的期望值(或试验数据)与各非随机变数之间的数学表达式为线性方程式，用最小二乘法对线性方程中的未知参数作出估计，然后对线性模型这个假设作显著性检验.

回归分析是数理统计学中最常用的统计方法. 例如在各种预报预测中，习惯上称预报对象这个随机变数为因变数，称与此有关的非随机变数为自变数(或预报因子)，需要建立因变数(或预报对象)与一组自变数(或一组预报因子)之间的数学关系式. 又如在寻求生产过程的工艺最优化问题中，先要寻求工艺的最优化区域，

在还不完全了解生产过程的物理、化学、生物等原理及经验的条件下，用回归分析来解决生产过程的工艺最优化问题是一个比较有效的数学方法.

对于诸如经验公式的求得、因子分析、产品的质量控制；水文、气象、虫害及地震等方面的预报；自动控制中的数学模型及某些新标准的制定等，回归分析往往是一种行之有效的数学工具.

试验设计是数理统计学的一个重要分支，它的主要内容是讨论如何合理地安排试验以及试验后的数据怎样作统计分析等. 系统地设计试验并考虑如何进行统计分析的概念，是由费歇在他的著作"试验之设计"中首先提出来的(参看[20]). 方差分析是试验结果的数据分析中一种常用的统计方法. 试验设计开始出现于农田科学实验，第二次世界大战后才逐渐地被应用于工业科学实验中，如今在各个领域中都有其广泛应用. 试验设计的内容很丰富，有拟因子设计，平衡不完全区组设计，正交设计等方面. 在这一章中，我们着重讨论数据分析中的方差分析法.

§9.1 线性模型

我们从几个例子谈起.

例9.1.1 我们知道，一个作等速直线运动的质点，在时刻 t 的位置是 s，一般它可以用公式 $s=\beta_0+\beta_1 t$ 来描写，其中 β_0 是质点在 $t=0$ 时刻的初始位置，β_1是平均速度. 如果 β_0 及β_1 是未知的，s 可以被观察，那么我们只要在两个不同的时刻观察质点的位置 s，就可以解出 β_0 及 β_1 的值. 由于种种原因，质点的位置不能被精确地观察，而是带有随机的测量误差. 也就是说 s 不能真正被观察，而观察到的数据是 $y=s+e$，其中 e 是测量(或观察)误差. 于是我们有公式

$$y=s+e=\beta_0+\beta_1 t+e, \tag{9.1.1}$$

其中 t 是可观察的非随机变数 对于 ε,我们把它看作是数学期望为零的随机变数,在一般情况下是允许这样做的,或者说这个观点点是符合实际情况的. 因此,y 看作为一个随机变数,它是可观察的随机变数,其数学期望为 s, β_0 及 β 是未知参数, 通过对 t 及 y 的 n 次观察得 n 组观察值 (t_i, y_i), $i=1,\cdots,n$, 即有

$$y_i=\beta_0+\beta_1 t_i+\varepsilon_i\ (i=1,\cdots,n).$$

下面用向量、矩阵的写法表示如下:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \tag{9.1.2}$$

其中

$$\mathbf{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad \boldsymbol{\varepsilon}=\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

$$\mathbf{X}=\begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, \quad \boldsymbol{\beta}=\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},$$

$$E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta}.$$

我们利用观察数据 $\mathbf{y}$ 及 $\mathbf{X}$ 对未知参数 $\boldsymbol{\beta}$ 作出估计.

例 9.1.2 小麦的亩产量记为 y, 它与水(x_1)、肥料(x_2)、土质(x_3)、麦种(x_4)、栽培技术(x_5)及管理措施(x_6)等因素有关. 就是说, 小麦亩产量 y 与 x_1, x_2, x_3, x_4, x_5 及 x_6 等都有一定的关系. 但是由于观察(或试验)中总存在随机因素的影响, 即使$x_1,\cdots x_6$相对固定,小麦的亩产量 y 也不完全相同,因此我们把 y 与 x_1, x_2, x_3, x_4, x_5 及x_6 的关系分为两部分来研究,即有

$$y=f(x_1,\cdots x_6)+\varepsilon, \tag{9.1.3}$$

其中 $f(x_1,\cdots,x_6)$表示 6 个能控制的因素 x_1, $\cdots$, x_6 与 y 的关系, $f(x_1\cdots, x_6)$是普通的函数表达式, 是非随机性的部分. ε 表示随机因素对小麦亩产量 y 的影响作用, 把 ε 看作数学期望为零的随

机变数．因此，y 为一随机变数，是可观察的随机变数，数学期望为 $f(x_1,\cdots x_6)$．在一般情况下，$f(x_1,\cdots x_6)$ 不一定是 $x_1,\cdots,x_6$ 的线性函数，为处理方便起见，可近似地当作线性的情形，即

$$y=\beta_1x_1+\cdots+\beta_6x_6+\varepsilon. \tag{9.1.4}$$

这样，(9.1.4)式是 y 关于因子 $x_1,\cdots,x_6$ 的线性函数，当我们对 $(y;x_1,\cdots,x_6)$ 进行 n 次观察(或试验)后，则(9.1.4)式是未知参数 $\beta_1,\cdots,\beta_6$ 的线性函数．

在这个例中，有些因子是非数量化的，如小麦种类是非数量化的因子．对于非数量化的因子，可适当给予数量化，以便于数学上计算分析．如小麦品种有甲、乙、丙三种，规定 x_4(甲)$=1$，x_4(乙)$=2$，x_4(丙)$=3$．在某些例子中，有些因子的取值过大而有些因子的取值过小，影响对因子所起作用大小的分析，这时可对各个因子的观察值作标准化处理，然后用标准化后的数值作回归分析计算．

例 9.1.3 在考察某化学反应时，发现其反应速度 y 主要决定于所用某种催化剂的用量 x 及所加的温度 T．当 x 及 T 改变时，反应速度 y 也跟着改变．但是由于观察(或测试)及其他条件的影响，当 x 及 T 固定时，所观察到的反应速度也不完全相同，即反应速度的观察值 y 与 x、T 之间存在着关系式

$$y=f(x,T)+\varepsilon. \tag{9.1.5}$$

其中 $f(x,T)$ 是 x 及 T 的普通二元函数，是非随机性的部分；ε 表示除 x 及 T 之外的对 y 起综合作用的各种次要条件，这些条件或是不可控制的，或是不可观察的．这些次要条件都归属于"事物之间总存在着随机因素的影响"这个范畴来考虑，我们把 ε 看作是数学期望为零的随机变数．因此，y 是数学期望为 $f(x,T)$ 的可观察的随机变数，即

$$E(y)=f(x,T).$$

我们通过试验(在 x 及 T 的不同的取值下进行试验)，能观察 y 的值，用回归分析法找出 $f(x,T)$ 的具体数学表达式．有了 $f(x,T)$

这个数学表达式，对今后实验工作很方便。例如，通过控制 x 及 T 以达到控制反应速度 y. 为处理方便起见，我们对(9.1.5)中的 $f(x,T)$ 用近似的线性函数来代替，即

$$y=\beta_0+\beta_1x+\beta_2T+e. \tag{9.1.6}$$

我们通过 n 组观察数据 $(y_i;x_i,T_i)$, $i=1,\cdots,n$, 对未知参数 β_0、β_1 及 β_2 作出估计.

综上所述，在这一章中，我们着重讨论线性回归分析，归纳出一般的线性模型.

设因变数 y 和自变数 $x_1,\cdots,x_k$ 之间服从如下线性关系：

$$y=\beta_1x_1+\cdots+\beta_kx_k+e. \tag{9.1.7}$$

对 y 及 $x_1,\cdots,x_k$ 同时作了 n 次观察(或试验)之后，得到 n 组数据 $(y_t;x_{t1},\cdots,x_{tk})$, $(t=1,\cdots,n)$. 它们满足下述线性关系式：

$$y_t=\beta_1x_{t1}+\cdots+\beta_kx_{tk}+e_t\ (t=1,\cdots,n) \tag{9.1.8}$$

(在这一章中我们用 t 表示下标). 用向量、矩阵形式记作

$$\mathbf{y}=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}_{n\times1},\qquad \boldsymbol{\beta}=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_k\end{pmatrix}_{k\times1}$$

$$\mathbf{X}=\begin{pmatrix}x_{11}\cdots x_{1k}\\ \vdots\qquad\vdots\\ x_{n1}\cdots x_{nk}\end{pmatrix}_{n\times k},\qquad \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix}_{n\times1},$$

则(9.1.8)式可表示为

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \tag{9.1.9}$$

其中 $\mathbf{X}$ 是已知的 $n\times k$ 阶常数矩阵；$\boldsymbol{\beta}$ 是 k 维的未知参数向量，$\boldsymbol{\varepsilon}$ 是数学期望为零的 n 维随机向量，$\mathbf{y}$ 是已知的 n 维观察向量.

我们对 $\boldsymbol{\varepsilon}$ 作如下假定：

$$\left.\begin{aligned}&E(\boldsymbol{\varepsilon})=\mathbf{0};\\&\operatorname{cov}(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}_{n\times n},\end{aligned}\right\}\tag{9.1.10}$$

其中 σ^2 是未知参数，$\mathbf{I}_{n\times n}$ 是单位矩阵，即有：

$$\left.\begin{aligned}&E(\varepsilon_t)=0;\\&D(\varepsilon_t)=\sigma^2;\\&\operatorname{cov}(\varepsilon_t,\varepsilon_s)=0,t\neq s\end{aligned}\right\}\tag{9.1.11}$$

$(t=1,\cdots,n,s=1,\cdots,n)$、就是说，对随机误差项 $\varepsilon_1,\cdots,\varepsilon_n$ 作这样的假定：无偏性、等方差性、不相关性，这种假定在一般情况下是合理的或允许的.

对于(9.1.9)中 $\mathbf{y}$，我们称它是服从线性模型，并且简记作 $(\mathbf{y},\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_{n\times n})$. 习惯上，我们说线性模型是指(9.1.9)及(9.1.10)两式组成. 由(9.1.9)及(9.1.10)两式容易算得

$$\left.\begin{aligned}&E(\mathbf{y})=\mathbf{X}\boldsymbol{\beta};\\&\operatorname{cov}(\mathbf{y},\mathbf{y})=\sigma^2\mathbf{I}_{n\times n}\end{aligned}\right\}\tag{9.1.12}$$

通常称(9.1.12)式为高斯-马尔科夫线性模型.

若对 ε_1, …, ε_n 作具有独立同分布 $N(0,\sigma)$的假定，则称(9.1.12)式为正态线性模型.

对于线性模型$(\mathbf{y},\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_{n\times n})$，所考虑的统计推断问题主要的是：

(1) 对未知参数向量 $\boldsymbol{\beta}$ 及 σ^2 进行估计；

(2) 对关于 $\boldsymbol{\beta}$ 的某种假设以及 $\mathbf{y}$ 服从线性模型的假设进行检验；

(3) 对 $\mathbf{y}$ 进行预测与控制.

在这一章中所作的讨论，总假定 $n>k$ 及常数矩阵 $\mathbf{L}$ 是满秩的，其中

$$\mathbf{L}=\mathbf{X}'\mathbf{X}①,$$

① 本书中矩阵 $\mathbf{A}$ 的转置用 $\mathbf{A}'$ 表示，例如此处以 $\mathbf{X}'$ 表示 $\mathbf{X}$ 的转置. 以下均同.

$\mathbf{L}$ 是 $k\times k$ 矩阵，它的秩 $\text{ran}(\mathbf{L})=k$. 关于 $\mathbf{L}$ 是降秩时的详细讨论,可参阅[21]的第 4 章.

§9.2 最小二乘法估计

这一段主要讨论线性模型(9.1.9)、(9.1.10)中未知参数 $\beta_1,\cdots,\beta_k$ 及 σ^2 的点估计,即 $(\mathbf{y},\mathbf{X\beta},\sigma^2\mathbf{I}_{n\times n})$ 中的未知参数向量 $\boldsymbol{\beta}$ 及 σ^2 的点估计.

一、参数的最小二乘法估计

我们考虑

$$Q=\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}=(\mathbf{y}-\mathbf{X\beta})'(\mathbf{y}-\mathbf{X\beta}), \tag{9.2.1}$$

亦即

$$Q=\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}=\sum_{t=1}^{n}e_t^2=\sum_{t=1}^{n}\left[y_t-\sum_{i=1}^{k}\beta_i x_{ti}\right]^2 \tag{9.2.2}$$

Q 是 n 次观察中误差项 e_t^2 之和，称 Q 为误差平方和，它反映了 y 与 $\sum_{i=1}^{k}\beta_i x_i$ 之间 n 次观察中总的误差程度.

最小二乘法原理是这样:寻找使得 Q 达到最小值的 $\hat{\beta}_1,\cdots,\hat{\beta}_k$ 作为 $\beta_1,\cdots,\beta_k$ 的点估计. 这时称 $\hat{\beta}_i$ 为参数 β_i 的最小二乘法估计量$(i=1,\cdots,k)$. 由于(9.2.2)中的 Q 是 β_i 的非负的二次函数，求 Q 的最小解是有可能的. 所要求的 $\hat{\beta}_i$ 应满足如下关系式:

$$\sum_{t=1}^{n}\left[y_t-\sum_{i=1}^{k}\hat{\beta}_i x_{ti}\right]^2=\min_{(\beta_1,\cdots,\beta_k)}\left\{\sum_{t=1}^{n}\left[y_t-\sum_{i=1}^{k}\beta_i x_{ti}\right]^2\right\}. \tag{9.2.3}$$

我们用微积分求满足(9.2.3)式的解 $\hat{\beta}_i$, $(i=1,\cdots,k)$. 将(9.2.2)式定义的 Q 对 β_i 分别求偏导数得:

$$\frac{\partial Q}{\partial \beta_i} = -2\sum_{t=1}^{n}\left[y_t - \sum_{j=1}^{k}\beta_j x_{tj}\right]\cdot x_{ti} \qquad i=1,\cdots,k.$$

令其偏导数等于零，方程组的解用 $\hat{\beta}_j$ 表示，即有

$$\sum_{t=1}^{n} y_t x_{ti} - \sum_{t=1}^{n}\sum_{j=1}^{k}\hat{\beta}_j x_{tj}\cdot x_{ti} = 0 \qquad i=1,\cdots,k,$$

或写为

$$\sum_{j=1}^{k}\hat{\beta}_j\left[\sum_{t=1}^{n} x_{tj}\cdot x_{ti}\right] = \sum_{t=1}^{n} y_t x_{ti} \qquad i=1,\cdots,k. \tag{9.2.4}$$

记

$$\sum_{t=1}^{n} x_{ti}x_{tj} = l_{ij}, \qquad \sum_{t=1}^{n} x_{ti}y_t = l_{iy},$$

$$\mathbf{L} = (l_{ij})_{k\times k} = \mathbf{X}'\mathbf{X}, \tag{9.2.5}$$

$\mathbf{L}$ 为对称的常数矩阵. 又记

$$\mathbf{X}'\mathbf{y} = (l_{iy})_{k\times 1}, \tag{9.2.6}$$

则(9.2.4)式可写为

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y},$$

即有

$$\mathbf{L}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{9.2.7}$$

通常称(9.2.7)式为正规方程，如果 $\mathbf{L}$ 是满秩的，即$|\mathbf{L}|\neq 0$，其秩为 k，那么正规方程(9.2.7)式有唯一解，记作

$$\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X}'\mathbf{y}, \tag{9.2.8}$$

其中 $\mathbf{L}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$ 为 $\mathbf{L}$ 的唯一的逆矩阵.

我们来证明(9.2.8)式中的 $\hat{\boldsymbol{\beta}}$ 为线性模型$(\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_{n\times n})$中未知参数向量 β 的最小二乘法估计量，亦即 $\hat{\beta}_i$ 满足关系式(9.2.

3)（$i=1,\cdots,k$）．现用矩阵形式来叙述其证明步骤．从(9.2.1)式知道，对任意 $\boldsymbol{\beta}$

$$Q=(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

则有

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})&=[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})+\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]'[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\\&\quad+\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]\\&=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\&\quad+(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\\&\geqslant(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}).\end{aligned}$$

上述证明过程中应用了如下结果：

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=[\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]'[\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})]\geqslant 0$$

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})&=(\mathbf{y}'\mathbf{X}-\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\&=(\mathbf{y}'\mathbf{X}-\mathbf{y}'\mathbf{X})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=0.\end{aligned}$$

至此，在 $|\mathbf{L}|\neq 0$ 时，证明了(9.2.8)式中的 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的最小二乘法估计量．

在实际工作中，常称

$$\hat{y}=\hat{\beta}_1x_1+\cdots+\hat{\beta}_kx_k \qquad (9.2.9)$$

为线性回归方程．在今后工作中，当自变数 x_1，$\cdots$，x_k 分别取值 $x_{1i},\cdots,x_{ki}$ 时，由(9.2.9)式算得 $\hat{y}_i$ 值作为 y_i 的估计值，即可用自变数 $x_1,\cdots,x_k$ 由(9.2.9)式来预测因变数 y（实际上是对随机变数 y 的数学期望 $E(y)$ 的估计）的取值．

二、最小二乘法估计量的性质

首先，我们对高斯-马尔科夫线性模型，讨论按(9.2.8)式确定 $\boldsymbol{\beta}$ 的最小二乘法估计量 $\hat{\boldsymbol{\beta}}$ 的一些基本性质．

性质 1 $\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{\beta}$ 的线性无偏估计量.

证 由于 $\hat{\boldsymbol{\beta}}=\mathbf{L}^{-1}\mathbf{X}'\mathbf{y}$，每一个 β_i 都是 $y_1,\cdots,y_n$ 的线性组合，因而 β_i 是 β_i 的线性估计量，即 $\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}$ 的线性估计量.

$$\begin{aligned}E(\hat{\boldsymbol{\beta}})&=E(\mathbf{L}^{-1}\mathbf{X}'\mathbf{y})=\mathbf{L}^{-1}\mathbf{X}'\cdot E(\mathbf{y})\\&=\mathbf{L}^{-1}\mathbf{X}'\cdot\mathbf{X}\boldsymbol{\beta}=\mathbf{L}^{-1}\mathbf{L}\boldsymbol{\beta}=\boldsymbol{\beta},\end{aligned}\tag{9.2.10}$$

即

$$E(\beta_i)=\beta_i\,(i=1,\cdots,k).\ \blacksquare$$

性质 2 $\hat{\boldsymbol{\beta}}$ 的协方差矩阵为 $\sigma^2\mathbf{L}^{-1}$

证 记 $\mathbf{B}=\mathbf{L}^{-1}\mathbf{X}'$，则 $\hat{\boldsymbol{\beta}}=\mathbf{B}\mathbf{y}$，

$$\begin{aligned}\operatorname{cov}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}})&=E\{[\mathbf{B}\mathbf{y}-\mathbf{B}E(\mathbf{y})][\mathbf{B}\mathbf{y}-\mathbf{B}E(\mathbf{y})]'\}\\&=\mathbf{B}\cdot E\{[\mathbf{y}-E(\mathbf{y})][\mathbf{y}-E(\mathbf{y})]'\}\cdot\mathbf{B}'\\&=\mathbf{B}\cdot\sigma^2\mathbf{I}_{n\times n}\cdot\mathbf{B}'\\&=\mathbf{L}^{-1}\mathbf{X}'\cdot\sigma^2\mathbf{I}_{n\times n}\cdot(\mathbf{L}^{-1}\mathbf{X}')'\\&=\mathbf{L}^{-1}\mathbf{X}'\cdot\sigma^2\mathbf{I}_{n\times n}\cdot\mathbf{X}\mathbf{L}^{-1}=\sigma^2\mathbf{L}^{-1}.\end{aligned}\tag{9.2.11}$$

记

$$\mathbf{L}^{-1}=\mathbf{C}=(c_{ij})_{k\times k},$$

则有

$$\left.\begin{aligned}&D(\beta_i)=\sigma^2\cdot c_{ii}\,(i=1,\cdots,k);\\&\operatorname{cov}(\beta_i,\beta_j)=\sigma^2c_{ij}\,(i,j=1,\cdots,k).\end{aligned}\right\}.\ \blacksquare$$

性质 3 记

$$\mathbf{e}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{y}-\hat{\mathbf{y}},\tag{9.2.12}$$

其中

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}},\tag{9.2.13}$$

称 $\mathbf{y}$ 为观察向量，$\hat{\mathbf{y}}$ 为由 $\mathbf{X}$ 对 $\mathbf{y}$ 的估计向量，$\mathbf{e}$ 为剩余向量. 则有：

$$E(\mathbf{e})=\mathbf{0}, \text{即 } E(e_t)=0,\ t=1,\cdots,n. \tag{9.2.14}$$

$$\mathrm{cov}(\hat{\boldsymbol{\beta}},\mathbf{e})=\mathbf{0}, \text{即对每一 } \hat{\beta}_i,\ i=1,\cdots,k,$$
$$\mathrm{cov}(\hat{\beta}_i,e_t)=0,\ t=1,\cdots,n. \tag{9.2.15}$$

证
$$\begin{aligned}E(\mathbf{e}) &= E(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})=E(\mathbf{y})-\mathbf{X}E(\hat{\boldsymbol{\beta}})\\ &=\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\boldsymbol{\beta}=\mathbf{0},\end{aligned}$$

$$\begin{aligned}\mathrm{cov}(\hat{\boldsymbol{\beta}},\hat{\mathbf{e}}) &= \mathrm{cov}(\hat{\boldsymbol{\beta}},\mathbf{y})-\mathrm{cov}(\hat{\boldsymbol{\beta}},\mathbf{X}\hat{\boldsymbol{\beta}})\\ &=\mathrm{cov}(\mathbf{L}^{-1}\mathbf{X}'\mathbf{y},\mathbf{y})-\mathrm{cov}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}})\mathbf{X}'\\ &=\mathbf{L}^{-1}\mathbf{X}'\cdot\sigma^2-\sigma^2\mathbf{L}^{-1}\mathbf{X}'=\mathbf{0}.\ \blacksquare\end{aligned}$$

最小二乘法估计的几何解释如下:

$$\begin{aligned}\mathrm{cov}(\mathbf{y},\mathbf{y}) &= \mathrm{cov}(\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e},\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{e})\\ &=\mathrm{cov}(\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\hat{\boldsymbol{\beta}})+\mathrm{cov}(\mathbf{e},\mathbf{e}),\end{aligned} \tag{9.2.16}$$

$$\begin{aligned}\mathbf{y}'\mathbf{y} &= (\hat{\mathbf{y}}+\mathbf{e})'(\hat{\mathbf{y}}+\mathbf{e})\\ &=\hat{\mathbf{y}}'\hat{\mathbf{y}}+\hat{\mathbf{y}}'\mathbf{e}+\mathbf{e}'\hat{\mathbf{y}}+\mathbf{e}'\mathbf{e}\\ &=\hat{\mathbf{y}}'\hat{\mathbf{y}}+\mathbf{e}'\mathbf{e},\end{aligned} \tag{9.2.17}$$

其中

$$\begin{aligned}\hat{\mathbf{y}}'\mathbf{e} &= \hat{\boldsymbol{\beta}}'\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\\ &=\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{y}-\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{0}.\end{aligned}$$

同理

$$\mathbf{e}'\hat{\mathbf{y}}=\mathbf{0}.$$

称$\sqrt{\mathbf{y}'\mathbf{y}}$、$\sqrt{\hat{\mathbf{y}}'\hat{\mathbf{y}}}$及$\sqrt{\mathbf{e}'\mathbf{e}}$分别为“观察向量”、“估计向量”及“剩余向量”的长度.

由正规方程(9.2.7)知道:

$$\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})=\mathbf{0},$$

即
$$\mathbf{X}'\mathbf{e}=\mathbf{0},$$

$$\sum_{t=1}^{n} x_{it} e_t = 0 (i=1, \cdots, k). \blacksquare$$

可见"剩余向量"e 与 k 个向量 $\mathbf{x}_i = (x_{1i}, \cdots, x_{ni})$ 都是正交的，$i=1, \cdots, k$. "估计向量"$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ 是属于 k 个向量 $\mathbf{x}_1, \cdots, \mathbf{x}_k$ 所张成的空间，这个空间称之为"估计空间". 可见，e 与"估计空间"垂直.

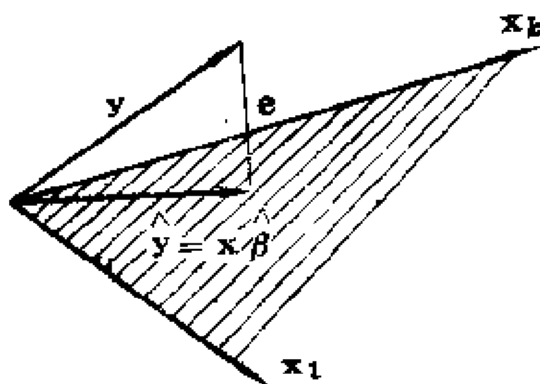


图 9.2.1

因此，我们可以这样说："在最小二乘法中，估计向量 $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ 是观察向量 y 在估计空间上的垂直投影".

性质 4 $E(\mathbf{e}'\mathbf{e}) = (n-k)\sigma^2$，记 $Q_e = \mathbf{e}'\mathbf{e}$，则 $E\left(\frac{Q_e}{n-k}\right) = \sigma^2$ 即 $\frac{Q_e}{n-k}$ 为 σ^2 的无偏估计量.

证 因为

$$Q_e = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{9.2.18}$$

$$\begin{aligned} EQ_e &= E(\mathbf{e}'\mathbf{e}) = E[\mathbf{y}'(\mathbf{I} - \mathbf{X}\mathbf{L}^{-1}\mathbf{X}')\mathbf{y}] \\ &= \mathbf{E}[(\mathbf{y} - E\mathbf{y})'(\mathbf{I} - \mathbf{X}\mathbf{L}^{-1}\mathbf{X}')(\mathbf{y} - E\mathbf{y})] \\ &\quad + (E\mathbf{y})'(\mathbf{I} - \mathbf{X}\mathbf{L}^{-1}\mathbf{X}')(E\mathbf{y}) \\ &= \sigma^2 \text{trace}[\mathbf{I} - \mathbf{X}\mathbf{L}^{-1}\mathbf{X}'] \\ &= n\sigma^2 - \sigma^2 \text{trace}(\mathbf{X}\mathbf{L}^{-1}\mathbf{X}') \\ &= n\sigma^2 - \sigma^2 \text{trace}(\mathbf{L}^{-1}\mathbf{X}\mathbf{X}') \\ &= n\sigma^2 - \sigma^2 \text{trace}(\mathbf{L}^{-1}\mathbf{L}) \\ &= n\sigma^2 - k\sigma^2 = (n-k)\sigma^2. \end{aligned}$$

其中 trace[A] 表示矩阵 A 的迹数. 记

$$\hat{\sigma}_e^2 = \frac{Q_e}{n-k}, \tag{9.2.19}$$

称 $\hat{\sigma}_e^2$ 为剩余方差，它是 σ^2 的无偏估计量. ∎

性质 5 $\hat{\boldsymbol{\beta}}$ 为 $\boldsymbol{\beta}$ 的最优线性无偏估计量.

证 $\hat{\boldsymbol{\beta}}=\mathbf{L}^{-1}\mathbf{X}'\mathbf{y}$, $E\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}$,

$$\operatorname{cov}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}})=\sigma^2\mathbf{L}^{-1},\ 0<\sigma^2<\infty.$$

设 $\mathbf{T}=\mathbf{A}\mathbf{y}$ 为 $\boldsymbol{\beta}$ 的任一线性无偏估计，则

$$E\mathbf{T}=\mathbf{A}\mathbf{X}\boldsymbol{\beta},$$

知

$$\mathbf{A}\mathbf{X}=\mathbf{I}_{k\times k},$$

$$\operatorname{cov}(\mathbf{T},\mathbf{T})=\mathbf{A}\operatorname{cov}(\mathbf{y},\mathbf{y})\mathbf{A}'=\sigma^2\mathbf{A}\mathbf{A}'. \tag{9.2.20}$$

记 $\mathbf{D}=\mathbf{A}\mathbf{A}'-\mathbf{L}^{-1}$，我们来证矩阵 $\mathbf{D}$ 是对角线元素为非负数的非负定矩阵. 事实上有：

$$\begin{aligned}&(\mathbf{A}-\mathbf{L}^{-1}\mathbf{X}')(\mathbf{A}-\mathbf{L}^{-1}\mathbf{X}')'\\&=\mathbf{A}\mathbf{A}'-\mathbf{A}\mathbf{X}\mathbf{L}^{-1}-\mathbf{L}^{-1}\mathbf{X}'\mathbf{A}'+\mathbf{L}^{-1}\mathbf{X}'\mathbf{X}\mathbf{L}^{-1}\\&=\mathbf{A}\mathbf{A}'-\mathbf{I}\mathbf{L}^{-1}-\mathbf{L}^{-1}\mathbf{I}+\mathbf{L}^{-1}\mathbf{L}\mathbf{L}^{-1}\\&=\mathbf{A}\mathbf{A}'-\mathbf{L}^{-1}=\mathbf{D}.\end{aligned}$$

这个性质通常称为高斯-马尔科夫定理. ∎

现在，我们对正态线性模型，进一步讨论 $\boldsymbol{\beta}$ 的最小二乘法估计量 $\hat{\boldsymbol{\beta}}$ 的分布函数问题，并讨论 Q_e 的分布函数问题. Q_e 是误差平方和 Q 在 $\boldsymbol{\beta}$ 用 $\hat{\boldsymbol{\beta}}$ 作估计量时达到的最小值，这可由(9.2.1)，(9.2.3)及(9.2.18)三式看到，今后也称 Q_e 为误差平方和.

性质 6 假定 $\varepsilon_1,\cdots,\varepsilon_n$ 为总体 ε 的子样，总体 ε 服从正态 $N(0,\sigma)$，则

(1) $\hat{\boldsymbol{\beta}}$ 与 $\mathbf{e}$ 相互独立；$\hat{\boldsymbol{\beta}}$ 与 Q_e 相互独立；

(2) $\hat{\boldsymbol{\beta}}$ 与 $\mathbf{e}$ 都服从正态分布，数学期望、协方差分别为

$$E(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta},\ \operatorname{cov}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}})=\sigma^2\mathbf{L}^{-1},$$

$$E(\mathbf{e})=0,\ \operatorname{cov}(\mathbf{e},\mathbf{e})=\sigma^2(\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}');$$

(3) $Q_e/\sigma^2 \sim \chi^2_{(n-k)}$. (9.2.21)

证 由(9.2.8)及(9.2.18)两式知,

$$\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X}'\mathbf{y},$$

$$Q_e = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'[\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}']\mathbf{y},$$

记

$$\mathbf{B} = \mathbf{L}^{-1}\mathbf{X}',\ \mathbf{A} = \mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}',$$

则

$$\hat{\boldsymbol{\beta}} = \mathbf{B}\mathbf{y},\quad Q_e = \mathbf{y}'\mathbf{A}\mathbf{y}.$$

(1) $\hat{\boldsymbol{\beta}}$ 为 $\mathbf{y}$ 的线性型(是指每一 β_i 为 $y_1,\cdots,y_n$ 的线性组合), Q_e 为 $\mathbf{y}$ 的半正定型(是指以 $y_1,\cdots,y_n$ 为变数的半正定型, 或称之为非负定型). $\mathbf{B}$ 为 $k\times n$ 矩阵,$\mathbf{A}$ 为对称的 $n\times n$ 矩阵. 从矩阵代数知道,存在一正交矩阵 $\mathbf{H}_{n\times n}$,使得 $\mathbf{D}=\mathbf{H}'\mathbf{A}\mathbf{H}$ 为对角线矩阵,记为

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n\times n},$$

其中 $\mathbf{D}_1$ 为 $n-k$ 阶的对角线元素非零的对角线矩阵. 同理,有

$$\mathbf{F} = \mathbf{B}\mathbf{H} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}_{k\times n}.$$

因为

$$\begin{aligned}\mathbf{B}\mathbf{A} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'-\mathbf{X}') = \mathbf{0},\end{aligned}$$

所以

$$\begin{aligned}\mathbf{0} &= \mathbf{B}\mathbf{A}\mathbf{H} = \mathbf{B}\mathbf{H}\mathbf{H}'\mathbf{A}\mathbf{H} = \mathbf{F}\mathbf{D} \\ &= \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.\end{aligned}$$

推得

$$\mathbf{F}_{11}\mathbf{D}_1 = \mathbf{F}_{21}\mathbf{D}_1 = \mathbf{0},\ 即\ \mathbf{F}_{11} = \mathbf{F}_{21} = \mathbf{0}.$$

因而

$$\mathbf{F}=\begin{pmatrix}\mathbf{0} & \mathbf{F}_{12}\\ \mathbf{0} & \mathbf{F}_{22}\end{pmatrix}=(\mathbf{0}\ \ \mathbf{F}_2),\qquad \mathbf{F}_2=\begin{pmatrix}\mathbf{F}_{12}\\ \mathbf{F}_{22}\end{pmatrix}.$$

$\mathbf{H}'$ 与 $\mathbf{H}$ 同为正交矩阵，今作正交线性变换

$$\mathbf{z}=\mathbf{H}'\mathbf{y}=\begin{pmatrix}\mathbf{z}_{(1)}\\ \mathbf{z}_{(2)}\end{pmatrix}. \tag{9.2.22}$$

于是 $z_1,\cdots,z_n$ 相互独立且皆服从正态分布. 由于

$$\mathbf{y}=\mathbf{H}\mathbf{z},\quad \mathbf{y}'=\mathbf{z}'\mathbf{H}',$$

于是

$$\left.\begin{aligned}&Q_e=\mathbf{y}'\mathbf{A}\mathbf{y}=\mathbf{z}'\mathbf{H}'\mathbf{A}\mathbf{H}\mathbf{z}=\mathbf{z}'\mathbf{D}\mathbf{z}=\mathbf{z}'_{(1)}\mathbf{D}_1\mathbf{z}_{(1)},\\ &\hat{\boldsymbol{\beta}}=\mathbf{B}\mathbf{y}=\mathbf{B}\mathbf{H}\mathbf{z}=\mathbf{F}\mathbf{z}=\mathbf{F}_2\mathbf{z}_{(2)},\end{aligned}\right\} \tag{9.2.23}$$

其中 $\mathbf{z}_{(1)}$ 是 $n-k$ 维正态随机向量，$\mathbf{z}_{(2)}$ 是 k 维正态随机向量，就是说，对于 $z_1,\cdots,z_n$，Q_e 是前 $n-k$ 个 z_i 的二次型，$\hat{\boldsymbol{\beta}}$ 是后 k 个 z_i 的线性型，因此 Q_e 与 $\hat{\boldsymbol{\beta}}$ 相互独立.

(2) $\hat{\boldsymbol{\beta}}$ 为 $\mathbf{y}$ 的线性型，$\mathbf{y}$ 为 n 维正态向量，则 $\hat{\boldsymbol{\beta}}$ 为 k 维正态向量. 但要注意，由于 $\operatorname{cov}(\beta_i,\beta_j)=\sigma^2c_{ij}$，所以 $\beta_1,\cdots,\beta_k$ 之间不一定相互独立.

由(9.2.12)式知

$$\mathbf{e}=\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}=\mathbf{y}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}'\mathbf{y}=(\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}')\mathbf{y},$$

所以 $\mathbf{e}$ 是 $\mathbf{y}$ 的线性型，是 n 维正态向量. 而且

$$\begin{aligned}\operatorname{cov}(\mathbf{e},\mathbf{e})&=\operatorname{cov}[(\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}')\mathbf{y},(\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}')\mathbf{y}]\\&=\sigma^2(\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}').\end{aligned}$$

(3) 记 $\mathbf{A}=\mathbf{I}_{n\times n}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}'$，则有

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})&=\mathbf{y}'\mathbf{A}\mathbf{y}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{y}\\&-\mathbf{y}'\mathbf{A}\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta}=\mathbf{y}'\mathbf{A}\mathbf{y},\end{aligned}$$

其中

$$\boldsymbol{\beta}'\mathbf{X}'\mathbf{A}=\boldsymbol{\beta}'\mathbf{X}'-\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\mathbf{L}^{-1}\mathbf{X}'=\boldsymbol{\beta}'\mathbf{X}'-\boldsymbol{\beta}'\mathbf{L}\mathbf{L}^{-1}\mathbf{X}'=0,$$

$$\mathbf{A}\mathbf{X}\boldsymbol{\beta}=\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\mathbf{L}^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=\mathbf{X}\boldsymbol{\beta}-\mathbf{X}\mathbf{L}^{-1}\mathbf{L}\boldsymbol{\beta}=0,$$

$$\boldsymbol{\beta}'\mathbf{X}\mathbf{A}\mathbf{X}\boldsymbol{\beta}=0.$$

于是

$$\frac{Q_e}{\sigma^2}=\frac{\mathbf{y}'\mathbf{A}\mathbf{y}}{\sigma^2}=\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})=\frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}}{\sigma^2}.$$

我们知道，矩阵 $\mathbf{A}$ 的秩为 $n-k$，可选正交矩阵 $\mathbf{H}$，使得 $\mathbf{D}=\mathbf{H}'\mathbf{A}\mathbf{H}$ 中的 $\mathbf{D}_1$ 为 $(n-k)\times(n-k)$ 阶矩阵. 又由 $\mathbf{H}$ 为正交矩阵知

$$\mathbf{H}'\mathbf{H}=\mathbf{H}\mathbf{H}'=\mathbf{I}_{n\times n},$$

于是

$$\frac{Q_e}{\sigma^2}=\frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}}{\sigma^2}=\left(\frac{\boldsymbol{\varepsilon}'}{\sigma}\right)\mathbf{H}'\mathbf{H}\mathbf{A}\mathbf{H}'\mathbf{H}\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)$$

$$=\mathbf{U}'\mathbf{H}\mathbf{A}\mathbf{H}'\mathbf{U}=\mathbf{U}'\mathbf{D}\mathbf{U},$$

其中

$$\mathbf{U}=\mathbf{H}\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right),\ \mathbf{U}'=\left(\frac{\boldsymbol{\varepsilon}'}{\sigma}\right)\mathbf{H}'.$$

$\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)$是每个分量服从正态 $N(0,1)$ 的随机向量. 因而

$\mathbf{U}'=(u_1,\cdots,u_n)$ 中每个向量也服从正态 $N(0,1)$，而且 $u_1,\cdots,u_n$ 相互独立. 所以

$$\frac{Q_e}{\sigma^2}=\mathbf{U}'\mathbf{D}\mathbf{U}$$

$$=\sum_{i=1}^{n-k}u_i^2\sim\chi^2_{(n-k)}.\ \blacksquare$$

最后顺便指出，最小二乘法估计量$\hat{\boldsymbol{\beta}}$ 及 $\hat{\sigma}^2=\frac{Q_e}{n}$，也是$\boldsymbol{\beta}$ 及 σ^2 的极大似然法估计量，因而 $\hat{\boldsymbol{\beta}}$ 及 $\hat{\sigma}^2$ 还有极大似然法估计量所具有的性质. 我们知道，$y_1,\cdots,y_n$ 是相互独立的随机变数，y_i 服从正

态 $N\left(\sum_{i=1}^{k}\beta_i x_{ti},\sigma\right)$, $t=1,\cdots,n$. 于是 $\beta_1,\cdots,\beta_k$ 及 σ^2 的似然函数为:

$$L(\boldsymbol{\beta},\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}\left(y_t-\sum_{i=1}^{k}\beta_i x_{ti}\right)^2\right],$$

可见，使得 $L(\beta,\sigma^2)$ 达到极大值是等价于使得(9.2.3)式达到极小值.

*§9.3 例 题

一、讨论三个例题.

例 9.3.1 我们考虑例 9.1.1 的表达式

$$y=\beta_0+\beta_1 x+\varepsilon, \tag{9.3.1}$$

$$\begin{cases} y_1=\beta_0+\beta_1 x_1+\varepsilon_1,\\ \quad\vdots \\ y_n=\beta_0+\beta_1 x_n+\varepsilon_n. \end{cases} \tag{9.3.2}$$

记
$$\mathbf{y}=\begin{pmatrix} y_1\\ \vdots\\ y_n\end{pmatrix},\quad \mathbf{X}=\begin{pmatrix} 1 & x_1\\ \vdots & \vdots\\ 1 & x_n\end{pmatrix},\quad \boldsymbol{\varepsilon}=\begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix},$$

则有

$$\mathbf{L}=\mathbf{X}'\mathbf{X}=\begin{pmatrix} n & n\bar{x}\\ n\bar{x} & \sum_{t=1}^{n}x_t^2\end{pmatrix},$$

$$\mathbf{L}^{-1}=\frac{1}{n\sum_{t=1}^{n}(x_t-\bar{x})^2}\begin{pmatrix}\sum_{t=1}^{n}x_t^2 & -n\bar{x}\\ -n\bar{x} & n\end{pmatrix},$$

其中 $\bar{x}=\frac{1}{n}\sum_{t=1}^{n}x_t$. 又记 $\bar{y}=\frac{1}{n}\sum_{t=1}^{n}y_t$, 因而

$$\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X}'\mathbf{y}$$

$$= \frac{1}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\begin{pmatrix} n\bar{y}\sum_{i=1}^{n}x_i^2 - n\bar{x}\sum_{i=1}^{n}x_i y_i \\ n\sum_{i=1}^{n}x_i y_i - n^2\bar{x}\bar{y} \end{pmatrix} \tag{9.3.3}$$

由此得到:

$$\beta_1 = \frac{n\sum_{i=1}^{n}x_i y_i - n^2\bar{x}\bar{y}}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2 - n\bar{x}^2} \tag{9.3.4}$$

由(9.3.2)可知

$$\beta_0 = \bar{y} - \beta_1\bar{x} \tag{9.3.5}$$

$$\sigma^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

$$= \frac{1}{n}\begin{pmatrix} y_1 - (\beta_0 + \beta_1 x_1) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_n) \end{pmatrix}'\begin{pmatrix} y_1 - (\beta_0 + \beta_1 x_1) \\ \vdots \\ y_n - (\beta_0 + \beta_1 x_n) \end{pmatrix}$$

$$= \frac{1}{n}\begin{pmatrix} (y_1 - \bar{y}) - \beta_1(x_1 - \bar{x}) \\ \vdots \\ (y_n - \bar{y}) - \beta_1(x_n - \bar{x}) \end{pmatrix}'\begin{pmatrix} (y_1 - \bar{y}) - \beta_1(x_1 - \bar{x}) \\ \vdots \\ (y_n - \bar{y}) - \beta_1(x_n - \bar{x}) \end{pmatrix}$$

$$= \frac{1}{n}\Big[\sum_{i=1}^{n}(y_i - \bar{y})^2 + \beta_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$- 2\beta_1\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\Big]$$

$$= \frac{1}{n}\Big[\sum_{i=1}^{n}(y_i - \bar{y})^2 - \beta_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2\Big]$$

$$= s_2^2 - \beta_1^2 s_1^2, \tag{9.3.6}$$

其中

$$s_1^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,\ s_2^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2.$$

则

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x$$

为 y 关于 x 的线性回归方程.

对于正态线性模型,则有

(1) $\hat{\beta}_0$ 服从正态 $N(\beta_0,\sigma_0)$, $\hat{\beta}_1$ 服从正态 $N(\beta_1,\sigma_1)$, 其中

$$\sigma_0^2=D(\hat{\beta}_0)=\sigma^2\sum_{i=1}^{n}x_i^2\Big/n\sum_{i=1}^{n}(x_i-\bar{x})^2=\sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right)\Big/ns_1^2,$$

$$\sigma_1^2=D(\hat{\beta}_1)=\sigma^2\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2=\sigma^2/ns_1^2,$$

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)=-\sigma^2\bar{x}\Big/\sum_{i=1}^{n}(x_i-\bar{x})^2=-\sigma^2\bar{x}/ns_1^2;$$

(2) $\frac{n\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{(n-2)}$, $E\left(\frac{n\hat{\sigma}^2}{n-2}\right)=\sigma^2$; 记 $\frac{n\hat{\sigma}^2}{n-2}=\hat{\sigma}_*^2$.

(3) $\hat{\beta}_0$ 与 $\hat{\sigma}^2$ 相互独立, $\hat{\beta}_1$ 与 $\hat{\sigma}^2$ 相互独立,于是

$$\frac{(\hat{\beta}_0-\beta_0)/\sigma_0}{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2}\Big/(n-2)}}=\sqrt{n-2}\,\frac{\hat{\beta}_0-\beta_0}{\hat{\sigma}\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\Big/s_1}\sim t_{(n-2)},\qquad(9.3.7)$$

$$\frac{(\hat{\beta}_1-\beta_1)/\sigma_1}{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2}\Big/(n-2)}}=\sqrt{n-2}\,\frac{\hat{\beta}_1-\beta_1}{\hat{\sigma}/s_1}\sim t_{(n-2)}.\qquad(9.3.8)$$

上式等价于

$$(n-2)\frac{(\hat{\beta}_1-\beta_1)^2}{\hat{\sigma}^2/s_1^2}=\frac{ns_1^2(\hat{\beta}_1-\beta_1)^2}{n\hat{\sigma}^2/(n-2)}=\frac{[(\hat{\beta}_1-\beta_1)/\sigma_1^2]^2}{\frac{n\hat{\sigma}^2}{\sigma^2}\Big/(n-2)}$$

$$\sim F_{(1,n-2)}.\qquad(9.3.9)$$

例 9.3.2 现考虑如下线性模型

$$y=\beta_0+\beta_1(x-\bar{x})+e, \tag{9.3.10}$$

$$\begin{cases} y_1=\beta_0+\beta_1(x_1-\bar{x})+e_1; \\ \quad\vdots \\ y_n=\beta_0+\beta_1(x_n-\bar{x})+e_n, \end{cases}$$

其中

$$\bar{x}=\frac{1}{n}\sum_{t=1}^{n}x_t.$$

这个模型的特点是对数据 $x_1,\cdots,x_n$ 作了中心化的处理. 先获得原来 n 组数据 (y_t,x_t),$(t=1,\cdots,n)$. 求出 $\bar{x}=\frac{1}{n}\sum_{t=1}^{n}x_t$,然后利用新的 n 组数据 $(y_t,x_t-\bar{x})$, $t=1,\cdots,n$, 建立经验线性回归方程.

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1(x-\bar{x}),$$

$$\mathbf{y}=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X}=\begin{pmatrix} 1 & x_1-\bar{x} \\ \vdots & \vdots \\ 1 & x_n-\bar{x} \end{pmatrix},$$

$$\mathbf{L}=\mathbf{X}'\mathbf{X}=\begin{pmatrix} n & 0 \\ 0 & \sum_{t=1}^{n}(x_t-\bar{x})^2 \end{pmatrix},$$

$$\mathbf{L}^{-1}=\frac{1}{n\sum_{t=1}^{n}(x_t-\bar{x})^2}\begin{pmatrix} \sum_{t=1}^{n}(x_t-\bar{x})^2 & 0 \\ 0 & n \end{pmatrix}.$$

记

$$s_1^2=\frac{1}{n}\sum_{t=1}^{n}(x_t-\bar{x})^2, \quad s_2^2=\frac{1}{n}\sum_{t=1}^{n}(y_t-\bar{y})^2,$$

因而

$$\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}\begin{pmatrix} n\bar{y}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\ n\sum_{i=1}^{n}(x_i-\bar{x})y_i \end{pmatrix}$$

$$= \frac{1}{ns_1^2}\begin{pmatrix} ns_1^2\bar{y} \\ ns_{12} \end{pmatrix}, \tag{9.3.11}$$

其中

$$\sum_{i=1}^{n}(x_i-\bar{x})y_i = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = ns_{12}.$$

由此得到

$$\left.\begin{aligned} &\hat{\beta}_0 = \bar{y}; \\ &\hat{\beta}_1 = \frac{s_{12}}{s_1^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \end{aligned}\right\} \tag{9.3.12}$$

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^{n}[y_i-\hat{y}_i]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}[(y_i-\bar{y})-\hat{\beta}_1(x_i-\bar{x})]^2 = s_2^2+\hat{\beta}_1^2 s_1^2-2\hat{\beta}_1 s_{12}$$

$$= s_2^2+\hat{\beta}_1^2 s_1^2-2\hat{\beta}_1^2 s_1^2 = s_2^2-\hat{\beta}_1^2 s_1^2. \tag{9.3.13}$$

记

$$R = \frac{s_{12}}{s_1 \cdot s_2}, \tag{9.3.14}$$

则有

$$\hat{\beta}_1 = \frac{s_{12}}{s_1^2} = R\frac{s_2}{s_1};$$

$$\hat{\sigma}^2 = s_2^2-\left(R\frac{s_2}{s_1}\right)^2 s_1^2 = (1-R^2)s_2^2. \tag{9.3.15}$$

对于正态线性模型，则有

(1) β_0 服从正态 $N(\beta_0, \sigma_0)$，β_1 服从正态 $N(\beta_1, \sigma_1)$，其中

$$\sigma_0^2=\frac{\sigma^2}{n},\quad \sigma_1^2=\frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sigma^2}{ns_1^2},$$

由 $\operatorname{cov}(\beta_0, \beta_1)=0$ 知，β_0 与 β_1 相互独立。

(2) β_0 与 $\hat{\sigma}^2$ 相互独立，β_1 与 $\hat{\sigma}^2$ 相互独立，

$$\frac{n\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{(n-2)},\quad \sqrt{n-2}\,\frac{\beta_0-\beta_0}{\hat{\sigma}}\sim t_{(n-2)}, \tag{9.3.16}$$

$$\sqrt{n-2}\,s_1\frac{\beta_1-\beta_1}{\hat{\sigma}}\sim t_{(n-2)}, \tag{9.3.17}$$

$$\frac{ns_1^2(\beta_1-\beta_1)^2}{n\hat{\sigma}^2/(n-2)}\sim F_{(1,n-2)}. \tag{9.3.18}$$

(3) 关于 y 的预测问题

由于 β_0 与 β_1 相互独立，则 $\hat{y}=\beta_0+\beta_1(x-\bar{x})$ 服从正态

$$N\left(\beta_0+\beta_1(x-\bar{x}),\ \sigma\sqrt{\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}\right),$$

于是 $y-\hat{y}$ 服从正态 $N\left(0,\ \sigma\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}\right)$，其中

$$D(\hat{y})=D(\beta_0)+(x-\bar{x})^2D(\beta_1)=\sigma^2\left[\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}\right], \tag{9.3.19}$$

$$\begin{aligned}D(y-\hat{y})&=D[(\beta_0-\beta_0)+(\beta_1-\beta_1)(x-\bar{x})+e]\\&=D(\beta_0-\beta_0)+(x-\bar{x})^2D(\beta_1-\beta_1)+D(e)\\&=\sigma^2\left[1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}\right].\end{aligned} \tag{9.3.20}$$

因此

$$\frac{(y-\hat{y})/\sqrt{D(y-\hat{y})}}{\sqrt{\frac{n\hat{\sigma}^2}{\sigma^2}/(n-2)}}=\frac{y-\hat{y}}{\hat{\sigma}_*\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}}\sim t_{(n-2)}, \tag{9.3.21}$$

其中 $\hat{\sigma}_e^2=\frac{n\hat{\sigma}^2}{n-2}$. 给定显著性水平 α，查得临界值 $t_{(n-2)}(\alpha)$，使得

$$P\left\{|y-\hat{y}|\leqslant\hat{\sigma}_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}t_{(n-2)}(\alpha)\right\}=1-\alpha, \tag{9.3.22}$$

亦即有

$$P\{\hat{y}_2\leqslant y\leqslant\hat{y}_1\}=1-\alpha, \tag{9.3.23}$$

其中

$$\hat{y}_1=\hat{\beta}_0+\hat{\beta}_1(x-\bar{x})+\hat{\sigma}_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}t_{(n-2)}(\alpha),$$

$$\hat{y}_2=\hat{\beta}_0+\hat{\beta}_1(x-\bar{x})-\hat{\sigma}_e\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}t_{(n-2)}(\alpha).$$

由上述两条曲线之间围成的区域，称之为 y 的置信水平为 $1-\alpha$ 的置信区域或预测区域.

例 9.3.3 炼钢是个氧化脱碳的过程. 因为钢液含碳量 x 与精炼时间 y 有一定的关系，现将某平炉共炼 34 炉钢记录下来的含碳量 x 与精炼时间 y 的数据列如表(9.3.1)，试求 y 与 x 的线性回归方程. 含碳量是指全部炉料熔化完毕时钢液的含碳量，精炼时间是指炉料熔毕至出钢所需要的时间.

解 我们按(9.3.10)式来确定线性回归方程. 当观察数据 (y_t, x_t), $(t=1,\cdots,n)$ 的数值比较大时，可通过下述变换而简化计算工作量. 令

$$x_t^*=\frac{x_t-c_1}{d_1},\ y_t^*=\frac{y_t-c_2}{d_2},\ t=1,\cdots,n,$$

其中 c_1、c_2 及 $d_1\neq0$、$d_2\neq0$ 为四个常数，则有:

$$x_t=c_1+d_1x_t^*,\ y_t=c_2+d_2y_t^*,\ t=1,\cdots,n,$$

$$\bar{x}=c_1+d_1\bar{x}^*,\ \bar{y}=c_2+d_2\bar{y}^*.$$

表 9.3.1

编 号	含碳量x(%)	精炼时间 y(分)	编号	含碳量x(%)	精炼时间 y(分)
1	1.80	200	18	1.16	100
2	1.04	100	19	1.23	110
3	1.34	135	20	1.51	180
4	1.41	125	21	1.10	130
5	2.04	235	22	1.08	110
6	1.50	170	23	1.58	130
7	1.20	125	24	1.07	115
8	1.51	135	25	1.80	240
9	1.47	155	26	1.27	135
10	1.45	165	27	1.15	120
11	1.41	135	28	1.91	205
12	1.44	160	29	1.90	220
13	1.90	190	30	1.53	[illegible]
14	1.90	210	31	1.55	160
15	1.61	145	32	1.77	185
16	1.65	195	33	1.77	205
17	1.54	150	34	1.43	160

其中

$$\bar{x}^*=\frac{1}{n}\sum_{i=1}^{n}x_i^*,\ \bar{y}^*=\frac{1}{n}\sum_{i=1}^{n}y_i^*,$$

并且

$$s_1^2=d_1^2 s_1^{*2},\ s_2^2=d_2^2 s_2^{*2},$$

其中

$$s_1^{*2}=\frac{1}{n}\sum_{i=1}^{n}(x_i^*-\bar{x}^*)^2,\ \ s_2^{*2}=\frac{1}{n}\sum_{i=1}^{n}(y_i^*-\bar{y}^*)^2.$$

由于

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\frac{d_1 d_2}{n}\sum_{i=1}^{n}(x_i^*-\bar{x}^*)(y_i^*-\bar{y}^*),$$

所以

$$R = R^*.$$

因此，由 (y_t^*, x_t^*), $t=1,\cdots,n$, 求出 $\bar{x}^*, \bar{y}^*, s_1^{*2}, s_2^{*2}, R^*$, 从而得到 $\bar{x}, \bar{y}, s_1^2, s_2^2$ 及 R.

实际计算如下(见表 9.3.2)

$$\sum_{t=1}^{34} x_t^* = 0.03, \bar{x}^* = 0.0009, 则\ \bar{x} = 1.5009;$$

$$\sum_{t=1}^{34} y_t^* = -12, \bar{y}^* = -0.353, 则\ \bar{y} = 158.235;$$

$$\sum_{t=1}^{34} (x_t^* - \bar{x}^*)^2 = \sum_{t=1}^{34} x_t^{*2} - 34(\bar{x}^*)^2 \approx 2.5463;$$

$$\sum_{t=1}^{34} (x_t^* - \bar{x}^*)(y_t^* - \bar{y}^*) = \sum_{t=1}^{34} x_t^* y_t^* - 34\bar{x}'\bar{y}' \approx 64.6508;$$

$$\sum_{t=1}^{34} (y_t^* - \bar{y}^*)^2 = \sum_{t=1}^{34} y_t^{*2} - 34(\bar{y}^*)^2 \approx 2003.76,$$

则

$$\sum_{t=1}^{34} (x_t - \bar{x})^2 \approx 2.5463, \sum_{t=1}^{34} (x_t - \bar{x})(y_t - \bar{y}) \approx 323.2540.$$

于是

$$\beta_0 = \bar{y} = 158.235, \ \beta_1 = \frac{323.2540}{2.5463} \approx 127.$$

从而得到精炼时间对含碳量的回归方程为:

$$\hat{y} = 158.235 + 127(x - 1.5009) = -32.38 + 127x. \qquad (9.3.24)$$

在本例中，取 $c_1 = 1.50, d_1 = 1, c_2 = 160, d_2 = 5$

从(9.3.24)看到，β_1 值很大，不必对 H_0: $\beta_1 = 0$ 作显著性检验.

若给定显著性水平 $\alpha = 0.05$, 按(9.3.23)式求出(9.3.10)式中

表 9.3.2

编 号	x^*	y^*	x^{*2}	y^{*2}	x^*y^*
1	0.30	8	0.0980	64	2.4
2	−0.46	−12	0.2116	144	5.52
3	−0.16	−5	0.0256	25	0.8
4	−0.09	−7	0.0081	49	0.63
5	0.54	15	0.2916	225	8.1
6	0.00	2	0.0000	4	0.0
7	−0.29	−7	0.0841	49	2.03
8	0.01	−5	0.0001	25	−0.05
9	−0.03	−1	0.0009	1	0.03
10	−0.05	1	0.0025	1	−0.05
11	−0.09	−5	0.0081	25	0.45
12	−0.06	0	0.0036	0	0.0
13	0.40	6	0.1600	36	2.4
14	0.40	10	0.1600	100	4.0
15	0.11	−3	0.0121	9	−0.33
16	0.15	7	0.0225	49	1.05
17	0.04	−2	0.0016	4	−0.08
18	−0.34	−12	0.1156	144	4.08
19	−0.27	−10	0.0729	100	2.7
20	0.01	4	0.0001	16	0.04
21	−0.40	−6	0.1600	36	2.4
22	−0.42	−10	0.1764	100	4.2
23	0.08	−6	0.0064	36	−0.48
24	−0.43	−9	0.1849	81	3.87
25	0.30	16	0.0900	256	4.8
26	−0.23	−5	0.0529	25	1.15
27	−0.35	−8	0.1225	64	2.8
28	0.41	9	0.1681	81	3.69
29	0.40	12	0.1600	144	4.8
30	0.03	−3	0.0009	9	−0.09
31	0.05	0	0.0025	0	0
32	0.27	5	0.0729	25	1.35
33	0.27	9	0.0729	81	2.43
34	−0.07	0	0.0049	0	0

y 的区间估计的端点．下面给出简化的近似计算公式(9.3.25)．

我们知道，$y-\hat{y}$ 服从正态 $N\left(0,\ \sigma\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}}\right)$，当 n 很大时且有下述关系式

$$1+\frac{1}{n}+\frac{(x-\bar{x})^2}{ns_1^2}\approx 1.$$

又 σ 用它的无偏估计量 $\hat{\sigma}_e=\sqrt{\frac{1}{n-2}Q_e}$ 作近似值，那么，

$$P\{|y-\hat{y}|\leqslant 2\hat{\sigma}_e\}\approx 0.95,$$

$$P\{\hat{y}-2\hat{\sigma}_e\leqslant y\leqslant \hat{y}+2\hat{\sigma}_e\}\approx 0.95. \qquad (9.3.25)$$

$$Q_e=\sum_{i=1}^{n}[y_i-\bar{y}]^2-U=ns_2^2-\hat{\beta}_1^2\cdot ns_1^2$$

$$=50094-41053=9041,$$

$\hat{\sigma}_e\approx 16.81$．于是，对于每一 x 值，y 的置信水平为 0.95 的预测区域为下述两条直线之间围成的区域：

$$\begin{cases} y=1.24+127x; \\ y=-66+127x. \end{cases} \qquad (9.3.26)$$

二、预测与控制

我们在例 9.3.2 与例 9.3.3 中，提到了线性回归方程中 y 的预测区域问题，它在实际工作中可用于处理线性预测与控制的问题．所谓预测问题，是指知道自变数 x 的取值状态，要估计 y 的取值范围．反之，如果需要把 y 的取值限制在某个范围，试问 x 的取值应该怎样控制，这是控制问题．如(9.3.21)～(9.3.23)式所作的讨论那样，同样可讨论更一般的线性模型(9.1.7)式中 y 的预测问题，即 y 的预测区域问题．

1．关于 y 的预测区域

对于(9.1.7)式中的因变数 y，在正态线性模型的基本假定

下，随机变数 y 服从正态 $N\left(\sum_{i=1}^{k}\beta_i x_i, \sigma\right)$，从性质 6 知道，$\hat{y}=\sum_{i=1}^{k}\hat{\beta}_i x$ 服从正态

$$N\left(\sum_{i=1}^{k}\beta_i x_i,\ \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}\sigma\right),$$

其中 c_{ij} 是矩阵 $\mathbf{C}=\mathbf{L}^{-1}$ 中的元素，而

$$\mathbf{L}=\mathbf{X}'\mathbf{X}=\left(\sum_{t=1}^{n}x_{ti}x_{tj}\right)_{k\times k}.$$

已知新的自变数$(x_1,\cdots,x_n)'$值，试求新的因变数 y 的区间估计．我们知道，y 与 $y_1,\cdots,y_n$ 相互独立，$\hat{y}$ 与 Q_e 相互独立，y 与 $\hat{y}$ 相互独立，y 与 Q_e 也相互独立，因此 $y-\hat{y}$ 与 Q_e 相互独立．记

$$z=y-\hat{y}=y-\sum_{i=1}^{k}\hat{\beta}_i x_i$$

则随机变数 z 服从正态分布，而且

$$E(z)=E(y-\hat{y})=E(y)-E(\hat{y})=\sum_{i=1}^{k}\beta_i x_i-\sum_{i=1}^{k}\beta_i x_i=0,$$

$$D(z)=D(y-\hat{y})=D(y)+D(\hat{y})=\sigma^2\left[1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j\right].$$

即随机变数 z 服从正态 $N\left(0,\ \sigma\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}\right)$

由 § 9.2 的性质 6 知，$Q_e/\sigma^2\sim\chi^2_{(n-k)}$．由于 z 与 Q_e 相互独立，所以

$$\frac{y-\hat{y}}{\sigma\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}}\Big/\sqrt{\frac{Q_e}{\sigma^2}\Big/(n-k)}.$$

$$
\frac{y-\hat{y}}{\hat{\sigma}_e\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}}\sim t_{(n-k)}. \tag{9.3.27}
$$

其中 $\hat{\sigma}_e=\sqrt{Q_e/(n-k)}$，$\hat{\sigma}_e^2$ 为剩余方差.

给定显著性水平 α，有临界值 $t_{(n-k)}(\alpha)$，使得

$$
P\{|t_{(n-k)}|\geqslant t_{(n-k)}(\alpha)\}=\alpha.
$$

因而

$$
P\left\{|y-\hat{y}|\leqslant t_{(n-k)}(\alpha)\cdot\hat{\sigma}_e\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}\right\}
$$

$$
=1-\alpha. \tag{9.3.28}
$$

也就是说，y 的置信水平为 $1-\alpha$ 的预测区域是由两个超越平面

$$
\hat{y}_1=\hat{y}+\hat{\sigma}_e t_{(n-k)}(\alpha)\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}
$$

与

$$
\hat{y}_2=\hat{y}-\hat{\sigma}_e t_{(n-k)}(\alpha)\sqrt{1+\sum_{i=1}^{k}\sum_{j=1}^{k}c_{ij}x_i x_j}
$$

之间所围成的区域，即有：

$$
P\{\hat{y}_2\leqslant y\leqslant\hat{y}_1\}=1-\alpha. \tag{9.3.29}
$$

2. 关于 β 的区间估计

记 $\theta=\sum_{i=1}^{k}a_i\beta_i$，其中 a_i 为常数，θ 为未知参数 β_i 的任一线性函数. 设 $\hat{\beta}_i$ 为 β_i 的最小二乘法估计量，则 θ 的最小二乘法估计量为：

$$
\hat{\theta}=\sum_{i=1}^{k}a_i\hat{\beta}_i. \tag{9.3.30}
$$

显然它是 θ 的线性无偏估计量，其方差为

$$D(\hat{\theta})=D\left(\sum_{i=1}^{k}a_i\hat{\beta}_i\right)=\sigma^2\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij},$$

其中 c_{ij} 为矩阵 $\mathbf{C}=\mathbf{L}^{-1}$ 中的元素，随机变数 $\hat{\theta}$ 服从正态

$$N\left(\sum_{i=1}^{k}a_i\beta_i,\sigma\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}}\right).$$

由于 $\hat{\theta}$ 与 Q_e 相互独立，所以

$$\frac{\hat{\theta}-\theta}{\sigma\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}}}\Bigg/\sqrt{\frac{Q_e}{\sigma^2}/(n-k)}$$

$$=\frac{\hat{\theta}-\theta}{\hat{\sigma}_e\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}}}\sim t_{(n-k)},$$

其中 $\hat{\sigma}_e=\sqrt{Q_e/(n-k)}$.

给定显著性水平 α，有临界值 $t_{(n-k)}(\alpha)$，使得

$$P\left\{\left|\frac{\hat{\theta}-\theta}{\hat{\sigma}_e\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}}}\right|\leqslant t_{(n-k)}(\alpha)\right\}=1-\alpha,$$

即

$$P\{\hat{\theta}_2\leqslant\theta\leqslant\hat{\theta}_1\}=1-\alpha. \tag{9.3.31}$$

其中

$$\hat{\theta}_1=\hat{\theta}+\hat{\sigma}_e t_{(n-k)}(\alpha)\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}},$$

$$\hat{\theta}_2=\hat{\theta}-\hat{\sigma}_e t_{(n-k)}(\alpha)\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{k}a_ia_jc_{ij}}.$$

特别地，当 $a_i=1$，且 $i\neq j$ 时 $a_j\neq 0, \theta=\beta_i (i=1,\cdots,k)$. 则有：

$$P\{\beta_{i2}\leqslant\beta_i\leqslant\beta_{i1}\}=1-\alpha. \tag{9.3.32}$$

其中

$$\left.\begin{aligned}\beta_{i1}&=\beta_i+\sigma_* t_{(n-k)}(\alpha)\sqrt{c_{ii}},\\ \beta_{i2}&=\beta_i-\sigma_* t_{(n-k)}(\alpha)\sqrt{c_{ii}},\end{aligned}\right\} \quad i=1,\cdots,k.$$

三、将曲线问题线性化

我们在例 9.3.1 至例9.3.3的三个例中，是先求出它们的经验线性回归方程，然后对“线性回归方程”作显著性检验. 其实，直观而又简便的办法是用 n 组观察数据(y_i,x_i)，$(i=1,\cdots,n)$，在平面上标出 n 个点，根据这 n 个点所呈现出的形状，我们粗略地看看象不象一条直线. 如果很不象一条直线，即使配了经验线性回归方程，线性模型的假设检验一定是作出被否定的判断. 我们可以根据观察数据的 n 点所呈现出形状，与常见的已知的函数图形作比较，选择一条曲线拟合这 n 个点，它比粗糙地配一条直线更为精确. 下面给出一些常见的、可以通过变数作变换而化成线性回归方程的函数图形及其数学表达式. 在化成线性回归方程之后，就可按最小二乘法估计其参数，从而给出原曲线方程中参数的估计.

1. 双曲线方程：

$$\frac{1}{y}=a+\frac{b}{x}.$$

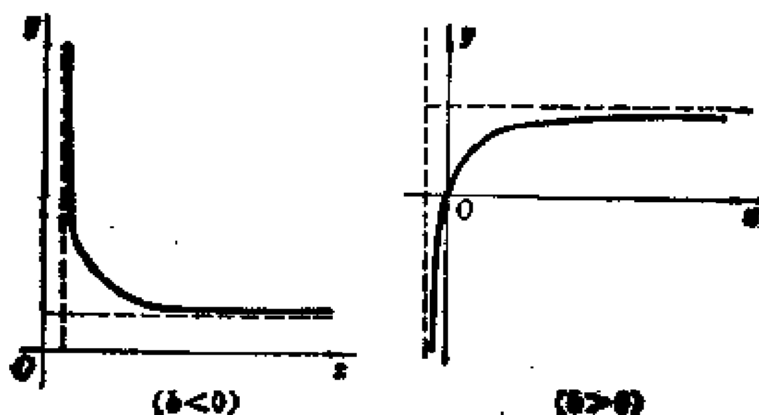


图 9.3.1

2. 幂函数方程:

$$y = ax^b.$$

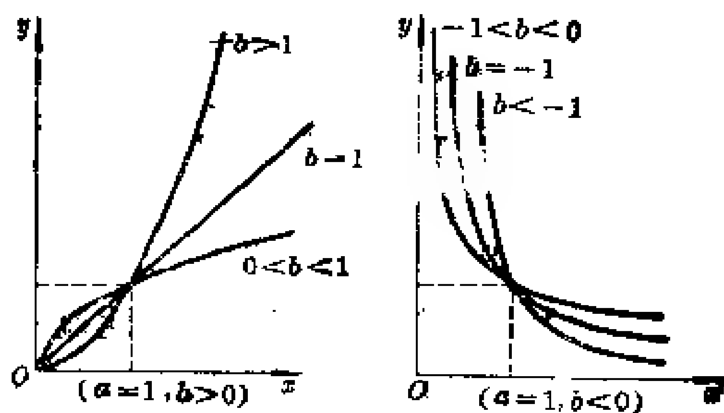


图 9.3.2

3. 指数曲线方程:

$$y = ae^{bx}.$$

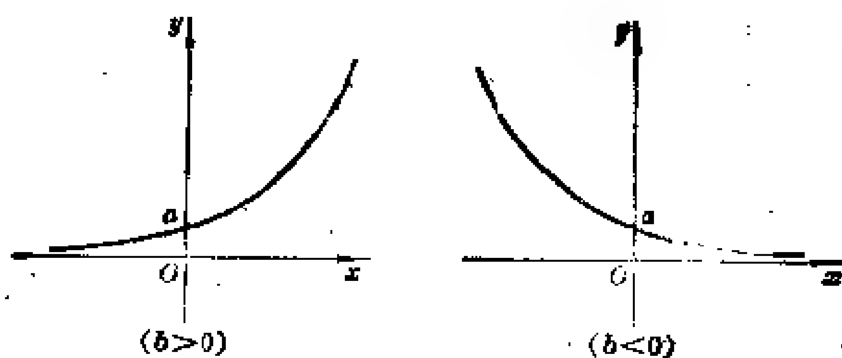


图 9.3.3

4. 指数曲线方程:

$$y = ae^{\frac{b}{x}}.$$

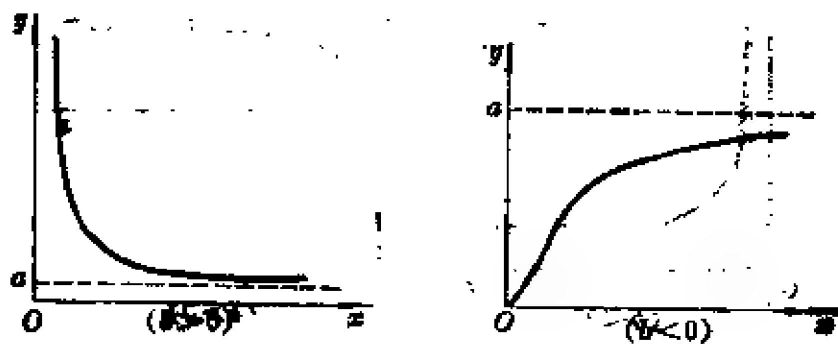


图 9.3.4

5. 对数曲线方程:

$$y=a+b\log x.$$

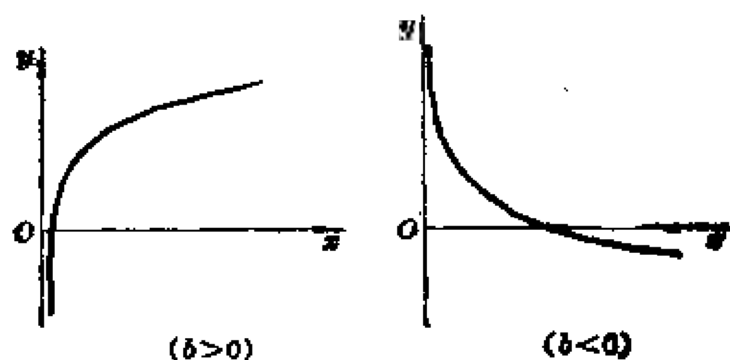


图 9.3.5

6. *S* 型曲线方程:

$$y=\frac{1}{a+be^{-x}}.$$

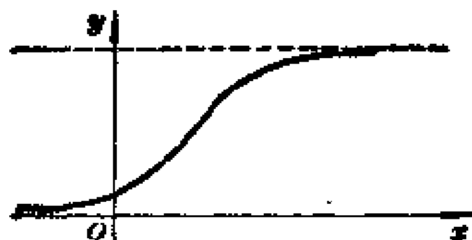


图 9.3.6

7. 抛物线方程:

$$y=b_0+b_1x+b_2x^2.$$

以上各曲线方程怎样通过变数作变换化成为线性方程，留作练习，在此从略．上述各曲线的图形，选自亨德一书①．

下面举两个例子．

例 9.3.4 在彩色显象中，根据以往的经验，形成染料光学密

① A. Hald. Statistical Theory With Engineering Applications. John Wiley & Sons, Inc. (1952).

表 9.3.3

编号	x	y	$x'=\frac{1}{x}$	$y'=\ln y$	x'^2	$x'y'$
1	0.05	0.10	20.000	−2.303	400.00	−46.06
2	0.06	0.14	16.667	−1.968	277.78	−32.77
3	0.07	0.23	14.286	−1.470	204.08	−21.00
4	0.10	0.37	10.000	−0.994	100.00	−9.94
5	0.14	0.59	7.143	−0.528	51.02	−3.77
6	0.20	0.79	5.000	−0.236	25.00	−1.18
7	0.25	1.00	4.000	0.000	16.00	0
8	0.31	1.12	3.226	0.113	10.41	0.36
9	0.38	1.19	2.652	0.174	6.93	0.46
10	0.43	1.25	2.326	0.223	5.41	0.52
11	0.47	1.29	2.128	0.255	4.53	0.54
$\sum_{t=1}^{11}$	2.46	8.07	87.408	−6.732	1101.16	−112.84

度 y 与析出银的光学密度 x 之间有下面类型的关系式:

$$y=ae^{-\frac{b}{x}},\quad b>0. \quad (9.3.33)$$

现对 y 及 x 同时作 11 次观察(或试验),获得 11 组数据 (y_t,x_t) $(t=1,\cdots,11)$,求出 a 及 b 的估计值. 将这 11 组数据 (y_t,x_t) 列在表 9.3.3 中,并将这 11 个点标在图 9.3.7 上.

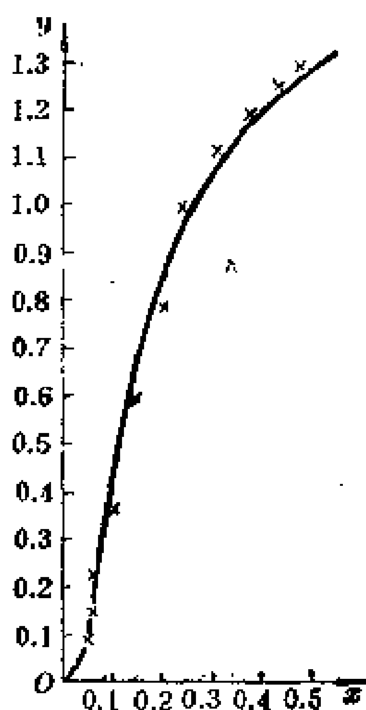


图 9.3.7

记

$$y'=\ln y,\quad x'=\frac{1}{x},$$

$$\beta_0=\ln a, \beta_1=-b,$$

则(9.3.33)式变换为:

$$y'=\beta_0+\beta_1 x'. \quad (9.3.34)$$

我们应用例 9.3.1 中 β_0 及 β_1 的计算公式，利用表 9.3.3 中的 y' 及 x' 的观察数据，在表 9.3.3 中计算 β_0 及 β_1 的估计值，可得：

$$\beta_0 = \bar{y}' - \beta_1 \bar{x}',$$

$$\beta_1 = \frac{\sum_{i=1}^{n} x'_i y'_i - n\bar{x}'\bar{y}'}{\sum_{i=1}^{n} x_i'^2 - n\bar{x}'^2}.$$

实际计算结果如下：

$$\bar{x}' = \frac{1}{11}\sum_{i=1} x'_i = \frac{87.408}{11} = 7.95,$$

$$\bar{y}' = \frac{1}{11}\sum_{i=1}^{11} y'_i = \frac{-6.732}{11} = -0.612,$$

$$\sum_{i=1}^{11} x'_i y'_i - 11\bar{x}'\bar{y}' = -112.84 + 53.49 = -59.35,$$

$$\sum_{i=1}^{11} x'^2 - 11\bar{x}'^2 = 1101.16 - 694.56 = 406.60,$$

$$\beta_1 = \frac{-59.35}{406.60} \approx -0.146,$$

$$\beta_0 = -0.612 - (-0.146) \times 7.95 \approx 0.549.$$

因而经验线性回归方程为

$$\hat{y}' = 0.549 - 0.146x'.$$

由于

$$\hat{a} = e^{\hat{\beta}_0} \approx 1.73,$$

$$\hat{b} = -\beta_1 \approx 0.146,$$

所以(9.3.33)式的经验回归曲线方程为

$$\hat{y} = 1.73e^{-\frac{0.146}{x}}. \qquad (9.3.35)$$

例 9.3.5 已知某种半成品在生产过程中的废品率 y 与它的某种化学成分 x 有关。现将观察（或试验）得到的一批数据，纪录在表 9.3.4 中，并将这 16 对数据 (x_i, y_i)，在平面上标出 16 个点，见图 9.3.8.

表 9.3.4

编号	y	$x_1=x$	$x_2=x^2$	编号	y	$x_1=x$	$x_2=x^2$
1	1.30	34	1156	9	0.44	40	1600
2	1.00	36	1296	10	0.56	41	1681
3	0.73	37	1369	11	0.30	42	1764
4	0.90	38	1444	12	0.42	43	1849
5	0.81	39	1521	13	0.35	43	1849
6	0.70	39	1521	14	0.40	45	2025
7	0.60	39	1521	15	0.41	47	2209
8	0.50	40	1600	16	0.60	48	2304

y 为废品率(%)，x 为化学成分含量(0.01%)。

从表 9.3.4 及图 9.3.8 中都可看到，废品率 y 最初随着化学成分 x 的增加而降低，而当化学成分 x 超过一定值之后，废品率 y 又有所回升．根据这个规律或散点分布图的形状，可以考虑拟合抛物线．

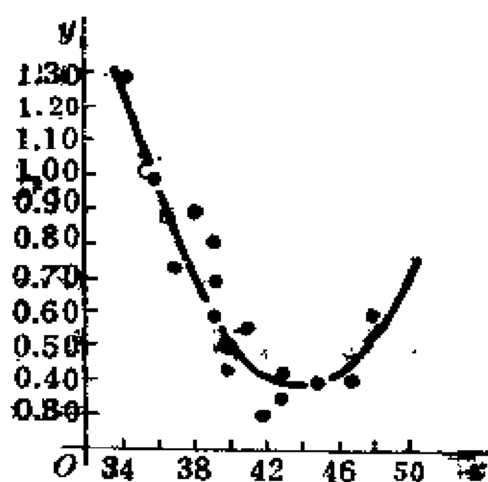


图 9.3.8

$$\hat{y}=\beta_0+\beta_1 x+\beta_2 x^2. \quad (9.3.36)$$

记 $\qquad x_1=x, \quad x_2=x^2,$

则(9.3.36)式变换为：

$$\hat{y}=\beta_0+\beta_1 x_1+\beta_2 x_2. \qquad (9.3.37)$$

下面我们利用表 9.3.4 中的 16 组数据 $(y_i; x_{1i}, x_{2i})$， $i=1,\cdots 16$，

像表 9.3.3 那样列表计算.

记

$$\mathbf{X}=\begin{pmatrix}1 & x_{11} & x_{12}\\ \vdots & \vdots & \vdots\\ 1 & x_{16,1} & x_{16,2}\end{pmatrix}_{16\times 3},$$

$$\mathbf{y}=\begin{pmatrix}y_1\\ \vdots\\ y_{16}\end{pmatrix}_{16\times 1},\qquad \boldsymbol{\beta}=\begin{pmatrix}\beta_0\\ \beta_1\\ \beta_2\end{pmatrix}_{3\times 1}.$$

由(9.2.7)及(9.2.8)两式知道,

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}=\mathbf{X}'\mathbf{y},\qquad \hat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

$$\mathbf{L}=\mathbf{X}'\mathbf{X}=\begin{pmatrix}16 & \sum_{t=1}^{16}x_{t1} & \sum_{t=1}^{16}x_{t2}\\ \sum_{t=1}^{16}x_{t1} & \sum_{t=1}^{16}x_{t1}^2 & \sum_{t=1}^{16}x_{t1}x_{t2}\\ \sum_{t=1}^{16}x_{t2} & \sum_{t=1}^{16}x_{t2}x_{t1} & \sum_{t=1}^{16}x_{t2}^2\end{pmatrix},$$

$$\mathbf{X}'\mathbf{y}=\begin{pmatrix}\sum_{t=1}^{16}y_t\\ \sum_{t=1}^{16}x_{t1}y_t\\ \sum_{t=1}^{16}x_{t2}y_t\end{pmatrix},$$

记

$$\bar{y}=\frac{1}{16}\sum_{t=1}^{16}y_t,\quad \bar{x}_1=\frac{1}{16}\sum_{t=1}^{16}x_{t1},\quad \bar{x}_2=\frac{1}{16}\sum_{t=1}^{16}x_{t2},$$

则有

$$\begin{cases}16\beta_0+\beta_1\cdot 16\bar{x}_1+\beta_2\cdot 16\bar{x}_2=16\bar{y};\\ 16\bar{x}_1\cdot\beta_0+\beta_1\sum_{t=1}^{16}x_{t1}^2+\beta_2\cdot\sum_{t=1}^{16}x_{t1}x_{t2}=\sum_{t=1}^{16}x_{t1}y_t;\\ 16\bar{x}_2\cdot\beta_0+\beta_1\sum_{t=1}^{16}x_{t2}x_{t1}+\beta_2\sum_{t=1}^{16}x_{t2}^2=\sum_{t=1}^{16}x_{t2}y_t,\end{cases}$$

其中

$$\beta_0=\bar{y}-\beta_1\bar{x}_1-\beta_2\bar{x}_2.$$

$$\begin{cases}\beta_1 l_{11}+\beta_2 l_{12}=l_{1y};\\ \beta_1 l_{21}+\beta_2 l_{22}=l_{2y}.\end{cases}$$

$$l_{iy}\hat{=}\sum_{t=1}^{16}x_{ti}y_t-16\bar{x}_i\cdot\bar{y},\qquad (i=1,2),$$

$$l_{ij}=\sum_{t=1}^{16}x_{ti}x_{tj}-16\bar{x}_i\cdot\bar{x}_j.\qquad (i,j=1,2).$$

实际计算结果如下:

$\bar{y}=0.6263,\quad \bar{x}_1=40.6875,$

$\bar{x}_2=1669.3125,$

$l_{11}=221.44,\quad l_{22}=1513685,$

$l_{12}=l_{21}=18283,$

$l_{1y}=-11.649,\quad l_{2y}=-923.05,$

$$\begin{cases}221.44\beta_1+18283\beta_2=-11.649;\\ 18283\beta_1+1513685\beta_2=-923.05.\end{cases}$$

由上式解得

$$\beta_1=-0.8205,\quad \beta_2=0.009301,$$

$$\beta_0=\bar{y}-\beta_1\bar{x}_1-\beta_2\bar{x}_2=18.484.$$

从而废品率 y 与化学成份 x 的关系式是:

$$\hat{y}=18.484-0.8205x+0.009301x^2, \tag{9.3.38}$$

它在

$$x_0=\frac{0.8205}{2\times0.009301}=44.11$$

时达到极小值，其极小值为

$$y_0=18.484-0.8205\times44.11+0.009301\times(44.11)^2$$
$$=0.39.$$

§9.4 假设检验与因子筛选

一、线性模型的假设检验

在实际工作中，事先我们并不能断定随机变数 y 与一组变数 $x_1,\cdots,x_k$ 之间确有线性关系．在求线性回归方程之前，线性模型(9.1.7)式只是一种假设，尽管这种假设常常不是没有根据的，但在求线性回归方程后，还是需要对这种线性回归方程同实际观察或试验数据拟合的效果进行检验．可提出原假设

$$H_0:\beta_1=\cdots=\beta_k=0. \tag{9.4.1}$$

如果 H_0 不否定，则表明用线性回归模型 (9.1.7) 不合适或拟合效果甚差．现在怎样建立统计量来对 H_0 作显著性检验呢？对于 $i=1,2,\cdots,k$ 中每一 i，可以对 $H_0:\beta_i=0$ 建立 F 检验统计量作显著性检验．

对于线性模型 (9.1.7)式，我们也可导出离差平方和的分解式，即有：

$$\begin{aligned}\mathbf{y}'\mathbf{y}&=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}+\mathbf{X}\hat{\boldsymbol\beta})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta}+\mathbf{X}\hat{\boldsymbol\beta})\\&=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol\beta})+\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol\beta}\\&=Q_e+U,\end{aligned} \tag{9.4.2}$$

其中 Q_e 为误差平方和，U 为回归平方和．现作验证：

$$U=\hat{\boldsymbol\beta}\mathbf{X}'\mathbf{X}\hat{\boldsymbol\beta}=[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]'\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]$$

$$=[\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'][\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}]$$

$$=\mathbf{y}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}, \tag{9.4.3}$$

$$Q_e=(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

$$=\mathbf{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}. \tag{9.4.4}$$

由此可见，将(9.4.3)与(9.4.4)两式相加，即可验证(9.4.2)式成立。

我们知道，y_t 服从正态 $N\left(\sum_{i=1}^{k}\beta_i x_{it},\sigma\right)$，$(t=1,\cdots,n)$，$y_1,\cdots,y_n$ 相互独立，当 H_0 成立时，那么 $y_1,\cdots,y_n$ 相互独立而且有相同分布 $N(0,\sigma)$，因而

$$\frac{1}{\sigma^2}\mathbf{y}'\mathbf{y}=\frac{1}{\sigma^2}\sum_{t=1}^{n}y_t^2\sim\chi_n^2. \tag{9.4.5}$$

由§9.2的性质6及定理6.3.3知

$$\frac{1}{\sigma^2}Q_e=\frac{n\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{(n-k)}, \tag{9.4.6}$$

$$\frac{1}{\sigma^2}U\sim\chi_k^2. \tag{9.4.7}$$

Q_e 与 U 相互独立，因而

$$\frac{U/k}{Q_e/(n-k)}=\frac{(n-k)U}{kQ_e}\sim F_{(k,n-k)}. \tag{9.4.8}$$

就是说，(9.4.8)式可作为对(9.4.1)式的 H_0 的检验统计量.

二、回归系数的假设检验

在线性回归模型(9.1.7)式中，并不意味着每个自变数 $x_1,\cdots x_k$ 对随机变数 y 的影响作用都是一样重要，可能有的起重要作用，而有的是可有可无的变数. 就是说，有的变数的系数 β 很大，而有的变数的系数 β 很小或为零. 因此，不否定线性模型之后，仍有必要从线性回归模型中，剔除那些次要的、可有可无的自

变数，只保留那些起重要作用的自变数，重新建立更为简练的线性回归方程，以更有利于实际工作．我们把自变数 x_i 叫作回归因子，其系数 β 称为 y 关于回归因子 x_i 的回归系数．我们说某个因子 x_i 对 y 的作用不显著，是指原假设

$$H_0: \beta_i = 0 \tag{9.4.9}$$

不否定．因此，对(9.4.9)式作显著性检验，即为挑选重要因子、剔除不显著因子的工作．

我们知道，β_i 服从正态 $N(\beta_i, \sigma\sqrt{c_{ii}})$，而且 $\hat{\beta}$ 与 Q_e 相互独立，因此当 H_0 成立时，则有

$$\frac{\beta_i^2/c_{ii}}{Q_e/(n-k)} = \frac{U_i/1}{Q_e/(n-k)} \sim F_{(1, n-k)}. \tag{9.4.10}$$

于是(9.4.10)式可作为对 H_0 作检验的检验统计量．记 $U_i = \frac{\beta_i^2}{c_{ii}}$，我们将指出 U_i 是 y 关于因子 x_i 的偏回归平方和．

若在线性回归方程

$$\hat{y} = \beta_1 x_1 + \cdots + \beta_k x_k \tag{9.4.11}$$

中剔除某个回归因子 $x_i (1 \leqslant i \leqslant k)$，这时重新建立的线性回归方程为：

$$\hat{y}^* = \beta_1^* x_1 + \cdots + \beta_{i-1}^* x_{i-1} + \beta_{i+1}^* x_{i+1} + \cdots + \beta_k^* x_k. \tag{9.4.12}$$

可用最小二乘法或极大似然法求出 y 关于 $x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_k$ 的回归系数 $\beta_1, \cdots, \beta_{i-1}, \beta_{i+1}, \cdots, \beta_k$ 的估计量 $\beta_1^*, \cdots, \beta_{i-1}^*, \beta_{i+1}^*, \cdots, \beta_k^*$. 一般地说，$\beta_i \neq \beta_i^*$．对于(9.4.12)式中的回归系数，再逐个作显著性检验，剔除不显著的回归因子，又重新建立 y 关于那些重要因子的新的线性回归方程．直至余下的回归因子都认为是显著的为止．

这样一来建立新的线性回归方程所需要的计算工作量是相当大的．但是，我们可以证明，剔除回归因子 x_i 后，(9.4.12)式中的

回归系数估计量 β_j^* 同(9.4.11)式中的回归系数估计量β_i 之间满足下述关系式

$$\beta_j^* = \beta_j - \frac{c_{ij}}{c_{ii}}\beta_i \qquad (i \neq j), \tag{9.4.13}$$

其中 c_{ij} 为 $\mathbf{C}=\mathbf{L}^{-1}=(\mathbf{X}'\mathbf{X})^{-1}$ 中的元素. 就是说，剔除一个因子 x_i 后，重新建立的线性回归方程的回归系数 $\beta_j^*(i\neq j)$，不必按最小二乘法重新来求，可直接利用前一方程的回归系数 β_i 由(9.4.13)式求得. (9.4.13)式成立的证明，留作练习.

我们知道，总离差平方和$\sum_{i=1}^{n}(y_i-\bar{y})^2$ 可以分解为两项 Q_e 和 U，其中的回归平方和U随回归因子的个数增加而增大. 因此，若在回归方程中剔除掉一个因子，则回归平方和U 只会减少而不会增加，减少的数值越大，说明该因子在回归方程中所起的作用越大，也就是说这个因子越重要、越显著.

设U 为 k 个因子所引起的回归平方和，U' 为剔除因子 x_i 后 $k-1$ 个因子所引起的回归平方和. 那么差

$$U_i = U - U' \tag{9.4.14}$$

就是去掉 x_i 后回归平方和减少的量，今后称 U_i 为因子 x_i 的偏回归平方和. 因此 U_i 的大小可以衡量因子 x_i 在回归方程中的作用大小，于是(9.4.14)式可以判断因子 x_i 是否显著. 可以证明:

$$U = \frac{1}{c_{ii}}\beta_i^2. \tag{9.4.15}$$

(9.4.15)式成立的证明，留作练习.

*三、最优回归的选择

上面所讨论的线性回归分析，是回归分析这个统计分支中的最基本的内容. 对回归分析的原理及应用要进一步作了解，可参看[20]及[27]等有关著作. 现在，我们粗略地讨论“最优” 回归方

程的选择问题．什么是最优回归方程呢？我们将以例 9.4.1为例作具体讨论。

在气象预报、地震预报、水文及虫害预报中，在自动控制生产过程的数学模拟中，我们常常从可能影响预报量 y 的许多因子中挑选若干个因子作为自变量，应用回归分析方法建立回归方程，以作预报或控制之用．问题是怎样在众多的因子中挑选那些对预报量 y 影响显著的自变量来建立回归方程，使得建立的回归方程中不包含对预报量 y 影响不显著的自变量，具有这种优点的回归方程，称之为“最优”回归方程．

我们知道，回归方程中包含的自变量个数越多，回归平方和 U 就越大，剩余平方和 Q_e 就越小．一般来讲，剩余平方和 Q_e 越小，预报的精确性越高．所以，“最优”回归方程中就希望包含尽可能多的自变量，特别是对预报量 y 有显著影响的自变量不能遗漏掉．但是，方程中含有的自变量个数太多，也有不利的一面．首先，计算复杂、使用不便．其次，如果回归方程中包含着对预报量 y 影响不显著的自变量，那么剩余平方和 Q_e 不会由于这些自变量的加入而减少多少，相反，由于 Q_e 的自由度 $(n-k)$ 随 k 增加而减少，反而使剩余方差 $\hat{\sigma}_e^2=Q_e/(n-k)$ 增大，降低预报的精确性．

例 9.4.1① 某种水泥在凝固时放出的热量 y 与水泥中下列4种化学成份有关：

x_1：$3CaO \cdot Al_2O_3$ 的成份(%)；

x_2：$3CaO \cdot SiO_2$ 的成份(%)；

x_3：$4CaO \cdot Al_2O_3 \cdot Fe_2O_3$ 的成份(%)；

---

① 此例的原始数据最早出处见于下文

H. Words, H. H. Steinour and H. R. Starke: "Effect of Composition of Portland Cement on Heat Evolved during Hardening". Industrial and Engineering Chemistry 24. 1932. 1207—14.

x_4: $2CaO \cdot SiO_2$ 的成份(%).

表 9.4.1 为测得的 13 组数据，试建立 y 关于这些因子的"最优"回归方程.

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4. \qquad (9.4.16)$$

表 9.4.1

编号	x_1	x_2	x_3	x_4	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

[方法 1] 从所有可能的因子组合的回归方程中挑选最优者.

在本例中，包含 1 个因子的回归方程共有 $C_4^1=4$个；包含 2 个因子的回归方程共有 $C_4^2=6$ 个；包含 3 个因子的回归方程共有 $C_4^3=4$ 个；包含 4 个因子的回归方程只有 1 个，共有 15 个回归方程. 对每个回归方程及其中的回归因子作显著性检验，然后挑选一个回归方程，其中所有的回归因子都显著，而且剩余方差较小.

现将各个回归方程的回归系数及其显著性、剩余方差列于表 9.4.2 中. 从这 15 个回归方程中，可以看出 $\hat{\sigma}_e^2$ 最小的为第(12)号方程，但其中含有不显著因子，因而它不是"最优"方程. 而全部

表 9.4.2

编号	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	Q_e	n_e	$\hat{\sigma}_e^2$
						2715.96	12	226.31
(1)	81.4793	1.8687^{**}				1265.69	11	115.06
(2)	57.4236		0.7891^{**}			906.34	11	82.39
(3)	110.2026			$-1.2558^{(*)}$		1939.40	11	176.31
(4)	117.5670				-0.7382^{**}	883.87	11	80.35
(5)	52.5773	1.4683^{**}	0.6623^{**}			57.90	10	5.79
(6)	72.3490	2.3125^{*}		0.4945		1227.02	10	122.71
(7)	108.0973	1.4400^{**}			-0.6140^{**}	74.76	10	7.48
(8)	92.0747		0.7313^{**}	-1.0084^{**}		415.44	10	41.54
(9)	94.1600		0.3109		-0.4569	868.88	10	86.89
(10)	131.2824			-1.1999^{**}	-0.7246^{**}	175.74	10	17.57
(11)	48.1396	1.6959^{**}	0.6530^{**}	0.2500		48.11	9	5.25
(12)	71.6482	1.4519^{**}	$0.4161^{(*)}$		-0.2365	47.97	9	5.33
(13)	111.6944	1.0519^{**}		$-0.4100^{(*)}$	-0.6428^{**}	50.84	9	5.65
(14)	203.6418		-0.9234^{**}	-1.4480^{**}	-1.5570^{**}	73.82	9	8.20
(15)	52.4052	$1.5511^{(*)}$	0.5101	0.1010	-0.1441	47.86	8	5.98

［说明］ 数字右上角注$^{(*)}$者为在 $\alpha=0.10$ 水平上显著，注*者为在 $\alpha=0.05$ 水平上显著，注**者为在 $\alpha=0.01$ 水平上显著. Q_e 为误差平方和，n_e 为其自由度，$\hat{\sigma}_e^2$ 为误差方差，是 σ^2 的无偏估计.

因子都显著且 $\hat{\sigma}_e^2$ 较小的为第(5)号回归方程，即

$$\hat{y}=52.5773+1.4683x_1+0.6623x_2 \qquad (9.4.17)$$

为"最优"回归方程.

这种方法当然总可以找到一个"最优"方程，然而工作量实在太大，实用上往往是行不通的. 例如有6个因子的话，就要建立 $(1+1)^6-1=63$ 个方程；有10个因子的话，就要建立 $2^{10}-1=1023$ 个方程.

［方法 2］ 从包含全部因子的回归方程中逐次剔除不显著因子.

在本例中首先建立第(15)号方程，然后对每个回归因子作显著性检验，剔除不显著因子中偏回归平方和最小的一个因子 x_3，重新建立 y 对 x_1, x_2, x_4 的回归方程，得第(12)号回归方程. 再对每个回归因子作显著性检验，剔除不显著因子 x_4，重新建立 y 对 x_1 及 x_2 的回归方程，得第(5)号回归方程. 由于第(5)号回归方程中所有因子都为显著因子，所以第(5)号为"最优"回归方程.

这种方法在因子不多且不显著因子也不多的情况下可以采用. 然而当因子较多时，特别是不显著因子很多时，计算工作量仍然是相当大的，因为每剔除一个因子后就得重新计算回归系数.

［方法 3］ 从一个因子开始，将回归因子逐个引入回归方程.

这种方法的要点是：先计算 y 同各个因子的相关系数，对于相关系数之绝对值最大的因子，对其偏回归平方和作显著性检验，如果显著就引入方程. 这种方法，只是对因子的引入把关；因子引入之后，不论以后是否会变成不显著性，概不剔除.

在本例中，先计算 y 同4个因子的各个相关系数，找到相关系数绝对值最大的一个因子 x_4，对其偏回归平方和 U_4 作显著性检验，经检验为显著，得第(4)号方程. 然后找出余下的因子中，与 y 的偏相关系数之绝对值最大的那个因子 x_1，同样作检验为显著，就

将因子 x_1 引入方程，得第(7)号方程. 再找余下因子中与 y 偏相关系数之绝对值最大的那个因子 x_2，经检验该因子也可引入，则得第(12)号方程. 最后，因子 x_3 经检验不显著，就不再引入了. 于是得到第(12)号回归方程. 这种方法计算工作量较小，然而这样得到的第(12)号方程，并不能保证其中所有引入的因子仍然是显著的，因为各个因子之间存在着相关关系. 在第(12)号回归方程中，发现引入的因子 x_4 变成不显著了，因而这个回归方程不是"最优"回归方程.

[方法 4]　逐步回归法

结合上述两种方法的特点，建立了逐步回归方法. 它的基本思想是：在所考虑的全部因子中，按其对预报量 y 作用的显著程度大小，挑选一个最重要的因子，建立只包含这个因子的回归方程；接着对其他因子计算偏回归平方和，引入一个显著性的因子，建立具有两个因子的回归方程. 从此之后，逐步回归的每一步(引入一个因子或从回归方程中剔除一个因子都算作一步）前后都要作显著性检验，即反复进行两个步骤：第一，对已在回归方程中的因子作显著性检验，显著者保留，把最不显著的那个因子从方程中剔除掉；第二，对不在回归方程中的其余因子，挑选最重要的那一个进入回归方程. 直到最后，回归方程中再也不能剔除任一因子，同时也不能再引入因子为止，保证最后所得的回归方程中所有因子都为显著因子.

逐步回归的数学模型与线性模型 (9.1.10)式及(9.1.9)式是一致的. 作为回归分析的内容来说，逐步回归并不包含多少新内容，基本过程是每步求解正规方程及计算偏回归平方和并作显著性检验. 主要技巧在于上述过程可以巧妙地通过一套计算程序以较少的计算量能够完成. 进一步的内容，可参看[27]的第三章.

§9.5 单因子方差分析

方差分析法是对试验(或观察)结果的数据作分析的一种常用的统计方法. 我们在第八章的§8.2中讲到的两个总体的数学期望是否相等的显著性检验,可以称之为单因子二水平的试验,在那里建立了t检验法. 在这一节中, 我们要讨论单因子多水平的试验, 将看到它在实际上是多个总体的数学期望是否相等的显著性检验. 对于这种假设检验问题, 在正态总体及方差相同的基本假定下,我们将建立F检验法. 这种F检验法,是由因子的方差同重复试验的误差项方差之比,建立F分布的检验统计量,因而又称作方差分析法.

一、基本思想

例 9.5.1 水稻品种比较试验

设水稻品种有珍珠矮、窄叶青、广进矮、朝阳早、南二矮等五种,现要考察水稻品种对水稻亩产量的影响作用,从中挑选水稻良种. 为此,可设水稻亩产量为指标,水稻品种为因子, 五个品种为该因子所取的五个水平,这是单因子五水平的试验.

例 9.5.2 灯丝的配料方案优选

某灯泡厂用四种不同配料方案制成的灯丝,生产了四批灯泡. 在每批灯泡中随机地抽取若干个灯泡测其使用寿命(单位:小时), 数据列于表 9.5.1 中.

表 9.5.1

灯丝 \ 使用寿命 \ 灯泡	1	2	3	4	5	6	7	8
甲	1600	1610	1650	1680	1700	1700	1780	
乙	1500	1640	1400	1700	1750			
丙	1640	1560	1600	1620	1640	1680	1740	1800
丁	1510	1520	1530	1570	1640	1680		

试问这四种灯丝生产的灯泡的使用寿命有无显著性的差异？在这里，可设灯泡的使用寿命时数为指标，灯泡的配料为因子，四种配料方案为四水平，这是单因子四水平的试验．如果由这四种配料方案制成的灯丝其灯泡的使用寿命没有显著性的差异，我们就可以从中选取一种较经济又方便的配料方案；如果有显著性的差异，则希望选取一种较优的配料方案，以对提高灯泡的使用寿命更为有利．

我们将以每一种配料方案制成灯丝后生产的灯泡视为一总体，灯泡使用寿命用ξ表示，则ξ为一随机变数．现用ξ_1, ξ_2, ξ_3及ξ_4分别表示这四批灯泡的使用寿命，即有四个总体：总体ξ_1，总体ξ_2，总体ξ_3及总体ξ_4．若从这四个总体中，分别随机抽取容量为n_i的子样$\xi_{i1}, \cdots, \xi_{in_i}(i=1, 2, 3, 4)$．我们将从这四个子样$\xi_{ij}$来推断这四个总体之间有无显著性差异．现假定总体$\xi_i$服从正态$N(a_i, \sigma)$，其中$a_i=E(\xi_i)(i=1, \cdots, 4)$．它们有相同的方差$\sigma^2$．因此，我们要讨论的问题就归结为判断原假设

$$H_0: a_1=a_2=a_3=a_4$$

是否成立，其中σ^2是未知参数．

由例 9.5.2 的讨论，将归纳出单因子多水平重复试验的数学模型及数据结构，如表 9.5.2 所示．

表 9.5.2

试验结果 重复 / 试验号	1	2	$\cdots$	n_i	平均	方差
1	ξ_{11}	ξ_{12}	$\cdots$	ξ_{1n_1}	$\bar{\xi}_1$	S_1^2
2	ξ_{21}	ξ_{22}	$\cdots$	ξ_{2n_2}	$\bar{\xi}_2$	S_2^2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
r	ξ_{r1}	ξ_{r2}		ξ_{rn}	$\bar{\xi}_r$	S_r^2

假定$\xi_1, \cdots \xi_r$相互独立，并假定ξ_i服从正态$N(a_i, \sigma)$，其中σ为

未知参数，$i=1,\cdots,r$. 现从总体 ξ_i 中随机抽取容量为 n_i 的子样 $\xi_{i1},\cdots,\xi_{in_i}$，$i=1,\cdots,r$.

我们考察如下原假设的检验问题：

$$H_0: a_1=\cdots=a_r. \tag{9.5.1}$$

记

$$\bar{\xi}_i=\frac{1}{n_i}\sum_{j=1}^{n}\xi_{ij},\qquad i=1,\cdots,r,$$

$$S_i^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)^2,\quad i=1,\cdots,r,$$

则 $\bar{\xi}_i$ 及 S_i^2 分别为第 i 个总体的子样平均及子样方差．记

$$\bar{\xi}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\xi_{ij}=\frac{1}{n}\sum_{i=1}^{r}n_i\bar{\xi}_i,$$

$$S^2=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi})^2,$$

其中 $n=\sum_{i=1}^{r}n_i$，称 $\bar{\xi}$ 及 S^2 分别为全体子样的平均及方差，简称作总平均及总方差.

下述恒等式成立：

$$\begin{aligned}
nS^2&=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi})^2\\
&=\sum_{i=1}^{r}\sum_{j=1}^{n_i}[(\xi_{ij}-\bar{\xi}_i)+(\bar{\xi}_i-\bar{\xi})]^2\\
&=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\xi_{ij}-\xi_i)^2+\sum_{i=1}^{r}\sum_{j=1}^{n}(\bar{\xi}_i-\bar{\xi})^2\\
&=\sum_{i=1}^{r}n_iS_i^2+\sum_{i=1}^{r}n_i(\bar{\xi}_i-\bar{\xi})^2,
\end{aligned}$$

其中

$$\sum_{j=1}^{n_i}(\xi_{ij}-\bar{\xi}_i)(\bar{\xi}_i-\bar{\xi})=0, i=1,\cdots r.$$

记

$$Q_e=\sum_{i=1}^{r}n_iS_i^2,\quad U_1=\sum_{i=1}^{r}n_i(\bar{\xi}_i-\bar{\xi})^2, \tag{9.5.2}$$

则有:

$$nS^2=Q_e+U_1. \tag{9.5.3}$$

记 $S_总^2=nS^2$,称 $S_总^2$ 为总离差平方和. 称(9.5.3) 式为总离差平方和的分解式,也称(9.5.2) 式中的 Q_e 为误差项平方和, U_1 为因子的平方和,在(9.5.2)式中, Q_e 是各总体中的子样方差 S_i^2 的线性组合,其中每个 S_i^2 都是 σ^2 的极大似然法估计量,是渐近无偏及渐近优效估计量,Q_e 反映了重复试验中误差的总大小. U_1 由各总体的子样平均 $\bar{\xi}_i$ 与总平均 $\bar{\xi}$ 之间的差异平方所组成,反映了各总体的子样平均之间的差异程度, 而子样平均 $\bar{\xi}_i$ 是 a_i 的无偏估计量且是优效估计量, 因此 U_1 在一定程度上可反映原假设 H_0 是否成立. 如果 U_1 显著地大于 Q_e, 说明各 $\bar{\xi}_i$ 之间的差异大小显著地大于重复试验中误差的总大小, 那么 H_0 可能不成立. 这种比较方差大小来判断原假设 H_0 是否成立的方法, 就是方差分析法的名称由来.

试问, U_1/Q_e 的比值大到什么程度, 可以否定 H_0 呢?我们由第六章中定理 6.3.3 的推论知道,当 H_0 成立时,$\bar{\xi}_1,\cdots,\bar{\xi}_r$ 是 r 个相互独立的且均服从正态$N(a,\sigma)$,因而统计量

$$\frac{U_1/(r-1)}{Q_e/(n-r)}\sim F_{(r-1,n-r)} \tag{9.5.4}$$

可以作为判断 H_0 是否成立的检验统计量.

给定显著性水平 α，怎样选择临界值呢？因为比值较大有可能否定 H_0，所以从右边取单边临界值．在 F 分布临界值表中，查第一自由度 $r-1$，第二自由度 $n-r$ 对应的临界值 $F_{(r-1,n-r)}(\alpha)$，使得

$$P\{F_{(r-1,n-r)}\geqslant F_{(r-1,n-r)}(\alpha)\}=\alpha \tag{9.5.5}$$

若按 (9.5.4) 式由子样观察值计算得 F 检验统计量的观察值 $F_{(r-1,n-r)}$ 大于 $F_{(r-1,n-r)}(\alpha)$，则在置信水平 $1-\alpha$ 下否定 H_0，在实际工作中称之为以 $1-\alpha$ 的把握认为这个因子对试验指标的影响是显著的，也就是 a_i 之间的差异是显著的．

表 9.5.3　方差分析表

项目	离差平方和	自由度	方差	F　值
因子 误差	U_1 Q_e	$r-1$ $n-r$	$U_1/(r-1)$ $Q_e/(n-r)$	$\dfrac{U_1/(r-1)}{Q_e/(n-r)}$
总和	S_T^2	$n-1$		

例 9.5.2 的实际计算结果见表 9.5.4．此时 $r=4,n=26$．

$F_{(3,22)}=2.15$.

表 9.5.4　方差分析表

方差来源	平方和	自由度	方　差	F　值
因子 误差	44361.2 151350.8	3 22	14787.1 6879.2	2.15
总和	195712	25		

给定显著性水平 $\alpha=0.05$，查得临界值 $F_{(3,22)}(0.05)=3.05$．因为 $2.15<3.05$，所以在置信水平 0.95 下不否定 H_0，也就是说这四种灯丝生产的灯泡其平均使用寿命之间没有显著性的差异．在实际工作中，通常说这个因子对试验指标的影响不显著．

*二、数学模型

现在，为了清楚地理解 Q_e 及 U_1 的统计意义，更好地了解各水平对试验指标的作用大小，现引进"效应"这个概念，并建立数学模型，以便进一步讨论多因子试验的方差分析.

考虑数据结构式：

$$\xi_{ij}=a_i+\varepsilon_{ij},\qquad \begin{array}{l} i=1,\cdots,r, \\ j=1,\cdots,n_i. \end{array} \tag{9.5.6}$$

对于数学模型(9.5.6)，作如下基本假定：总体 ξ_i 服从正态 $N(a_i,\sigma)$，$a_i=E(\xi_i)$ $(i=1,\cdots,r)$，具有相同的方差 σ^2 且 $\xi_1,\cdots,\xi_r$ 相互独立，$\xi_{i1},\cdots,\xi_{in_i}$ 为总体 ξ_i 的容量为 n_i 的子样. ξ_{ij}为因子的第 i 个水平(即第 i 个总体)下第 j 次重复试验的试验结果的数据（即试验指标的观察值），ε_{ij} 为其相应的试验误差. a_i 是第 i 个总体 ξ_i 的数学期望，其实际意义是因子第 i 个水平下试验结果数据的理论均值.

在上述基本假定下，对于每一个总体ξ_i，即在因子的每一个水平 ξ_i 下，试验误差 ε_{i1}，$\cdots\varepsilon_{in_i}$ 是相互独立的且都服从正态 $N(0,\sigma)$的随机变数.

在数学模型(9.5.6)式中，记

$$\bar{\varepsilon}_i=\frac{1}{n}\sum_{j=1}^{n_i}\varepsilon_{ij}(i=1,\cdots,r),$$

$$\bar{\varepsilon}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\varepsilon_{ij}=\frac{1}{n}\sum_{i=1}^{r}n_i\bar{\varepsilon}_i.$$

在上述基本假定下，则有

$$E(\bar{\varepsilon}_i)=0,\ (i=1,\cdots r),$$

$$E(\bar{\varepsilon}_i-\bar{\varepsilon})=0\quad(i=1,\cdots,r),$$

$$E(\bar{\varepsilon})=0,$$

$$E\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}(e_{ij}-\bar{e}_i)^2\right]=\sum_{i=1}^{r}E\left[\sum_{j=1}^{n_i}(e_{ij}-\bar{e}_i)^2\right]$$

$$=\sum_{i=1}^{r}(n_i-1)\sigma^2=(n-r)\sigma^2, \qquad (9.5.7)$$

$$E\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}(e_{ij}-\bar{e})^2\right]=(n-1)\sigma^2. \qquad (9.5.8)$$

由(9.5.3)式知

$$E\left[\sum_{i=1}^{r}n_i(\bar{e}_i-\bar{e})^2\right]=E\left[\sum_{i=1}^{r}\sum_{j=1}^{n_i}(e_{ij}-\bar{e})^2-\sum_{i=1}^{r}\sum_{j=1}^{n_i}(e_{ij}-\bar{e}_i)^2\right].$$

$$=(n-1)\sigma^2-(n-r)\sigma^2=(r-1)\sigma^2. \qquad (9.5.9)$$

在数学模型(9.5.6)式中，记

$$a=\frac{1}{n}\sum_{i=1}^{r}n_ia_i$$

$$c_i=a_i-a \quad (i=1,\cdots,r).$$

称 c_i 为第 i 水平对试验指标的效应值，它反映因子第 i 水平对试验指标的“纯”作用大小. 可见，a_i 之间的差异性同 c_i 之间的差异性是等价的，并有

$$\sum_{i=1}^{r}n_ic_i=0.$$

于是，数学模型(9.5.6)式可以写成效应分解式：

$$\xi_{ij}=a+c_i+e_{ij}, \quad \begin{matrix} i=1,\cdots,r, \\ j=1,\cdots,n_i. \end{matrix} \qquad (9.5.10)$$

并有

$$\bar{\xi}_i=a_i+\bar{e}_i \quad (i=1,\cdots,r),$$

$$\bar{\xi}=a+\bar{e}.$$

*三、统计分析

对于数学模型(9.5.10),统计分析的任务归纳为三个方面:

(1) 求出 c_i 的点估计($i=1,\cdots,r$);

(2) 判断 $H_0: c_1=\cdots=c_r$;

(3) 求出 a_i 的区间估计.

现就上述三个问题逐一作讨论.

(1) 我们知道,$\bar{\xi}_i$ 是 a_i 的无偏估计($i=1,\cdots,r$),而因为

$$E(\bar{\xi})=E\left(\frac{1}{n}\sum_{i=1}^{r}n_i\bar{\xi}_i\right)=\frac{1}{n}\sum_{i=1}^{r}n_ia_i=a.$$

所以 $\bar{\xi}$ 为 a 的无偏估计. 记

$$\hat{c}_i=\bar{\xi}_i-\bar{\xi}\quad(i=1,\cdots,r),$$

由于

$$E(\hat{c}_i)\quad E(\bar{\xi}_i-\bar{\xi})=a_i-a=c_i\quad(i=1,\cdots r),$$

所以 $\hat{c}_i$ 是效应 c_i 的无偏估计.

(2) 现进一步说明(9.5.2)式中统计量 U_1 及 Q_e 的统计意义,易见,

$$E(Q_e)=E\left(\sum_{i=1}^{r}n_iS_i^2\right)=\sum_{i=1}^{r}E(n_iS_i^2)$$

$$=\sum_{i=1}^{r}(n_i-1)\sigma^2=(n-r)\sigma^2.$$

记 $S_e^2=Q_e/(n-r)$,则有

$$E(S_e^2)=\sigma^2.\qquad(9.5.11)$$

不论因子取什么水平,S_e^2 都是 σ^2 的无偏估计,Q_e 反映了重复试验的试验误差大小,称 S_e^2 为重复试验的误差项方差. 由于

$$(\bar{\xi}_i-\bar{\xi})^2=[(a_i+\bar{\varepsilon}_i)-(a+\bar{\varepsilon})]^2$$

$$=(\bar{\varepsilon}_i-\bar{\varepsilon})^2+2(\bar{\varepsilon}_i-\bar{\varepsilon})(a_i-a)+(a_i-a)^2,$$

所以

$$E(U_1)=E\left[\sum_{i=1}^{r}n_i(\bar{\xi}_i-\bar{\xi})^2\right]$$

$$=\sum_{i=1}^{r}E[n_i(\bar{\varepsilon}_i-\bar{\varepsilon})^2]+\sum_{i=1}^{r}E[n_i(a_i-a)^2]$$

$$+2\sum_{i=1}^{r}n_i(a_i-a)\cdot E(\bar{\varepsilon}_i-\bar{\varepsilon})$$

$$=(r-1)\sigma^2+\sum_{i=1}^{r}n_ic_i^2.$$

可见，U_1 也反映了试验误差的大小，而且主要的是同各水平的效应 c_i 有关，即同这个因子所取的水平有关. 记 $S_{因子}^2=U_1/(r-1)$，则有

$$E(S_{因子}^2)=\sigma^2+\frac{1}{r-1}\sum_{i=1}^{r}n_ic_i^2. \qquad (9.5.12)$$

当原假设 H_0 成立时，$S_{因子}^2$ 才是 σ^2 的无偏估计量. 称 $S_{因子}^2$ 为因子的方差.

当原假设 H_0 成立时，由(9.5.4)式知道，统计量

$$F=\frac{S_{因子}^2}{S_e^2}=\frac{U_1/(r-1)}{Q_e/(n-r)}\sim F_{(r-1,\,n-r)}, \qquad (9.5.13)$$

因而它可作为判断 H_0 是否成立的检验统计量. 在实际工作中，称这个统计量的观察值为因子的 F 值.

(3) 求各 a_i 的区间估计

若原假设 H_0 被否定，则因子的各水平的效应 c_i 之间有显著性的差异，从而挑选效应最大的水平，称之为优水平. 在确定了因子的优水平之后，进一步要问在这个优水平下试验指标的观察值怎样预测？这就要求出各 a_i 的区间估计.

我们知道，对于每一总体 ξ_i，若 ξ_i 服从正态 $N(a_i,\sigma)$，则子样

平均 $\bar{\xi}_i$ 与子样方差 S_i^2 相互独立；又若 $\xi_1,\cdots,\xi_r$ 相互独立，利用第六章的 § 6.3 中的三个定理及其推论知，统计量

$$\left(\frac{\bar{\xi}_i - a_i}{\sigma/\sqrt{n_i}}\right)^2 \sim \chi_1^2;$$

$$\frac{Q_e}{\sigma^2} = \sum_{i=1}^{r} \frac{n_i S_i^2}{\sigma^2} \sim \sum_{i=1}^{r} \chi_{(n_i-1)} = \chi^2_{(n-r)}.$$

所以，统计量

$$\frac{(\bar{\xi}_i - a_i)^2 \Big/ \frac{\sigma^2}{n_i}}{\frac{Q_e}{\sigma^2} \Big/ (n-r)} = \frac{n_i(\bar{\xi}_i - a_i)^2}{S_e^2} \sim F_{(1,\, n-r)}.$$

给定显著性水平 α，查表得临界值 $F_{(1,\, n-r)}(\alpha)$，使得

$$P\left\{\frac{n_i(\bar{\xi}_i - a_i)^2}{S_e^2} \geqslant F_{(1,\, n-r)}(\alpha)\right\} = \alpha, \qquad (9.5.14)$$

因而有

$$P\left\{\bar{\xi}_i - \frac{S_e}{\sqrt{n_i}}\sqrt{F_{(1,\, n-r)}(\alpha)} < a_i < \bar{\xi}_i + \frac{S_e}{\sqrt{n_i}}\sqrt{F_{(1,\, n-r)}(\alpha)}\right\}$$

$$= 1 - \alpha. \qquad (9.5.15)$$

即得 a_i 的置信水平为 $1-\alpha$ 的区间估计，$i = 1, \cdots, r$.

习　　题

1. 今有 10 组观测数据由下表给出：

x	0.5	−0.8	0.9	−2.8	6.5	2.3
y	−0.3	−1.2	1.1	−3.5	4.6	1.8
x	1.6	5.1	−1.9	−1.5		
y	0.5	3.8	−2.8	0.5		

应用线性模型：

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

假定误差项 e 服从正态 $N(0,\sigma)$，$e_1,\cdots,e_{10}$ 为其子样.

(1) 求 β_0 及 β_1 的最小二乘法估计;

(2) 求 β_1 的置信水平为 0.95 的区间估计;

(3) 在显著性水平 $\alpha=0.05$ 下检验假设 $H_0:\beta_1=0$;

(4) 计算剩余方差 $\hat{\sigma}_e^2$;

(5) 求 $y=\beta_0+\beta_1 x$ 的置信水平为 0.95 的预测区域.

2. 在维尼纶醛化试验中，固定其它因素，考虑醛浓度与反应时间对醛化度的关系，试验数据如下:

醛化度 \ 反应时间 / 甲醛	3	5	7	12	20	30
32.10	17.8	22.9	25.9	29.9	32.9	35.4
33.00	18.2	22.9	25.1	28.6	31.2	34.1
27.60	16.8	20.0	23.6	28.0	30.0	33.1

记醛化度为 y，反应时间为 x_1，甲醛浓度为 x_2，由经验知道，y 与 x_2 成正比，而与 x_1 成反比，并有

$$\hat{y}=b_0+b_1\frac{1}{x_1}+b_2x_2.$$

试求 b_0、b_1 及 b_2 的最小二乘法估计值.

3. 物体降落的距离 s 与时间 t 的关系，由下式确定

$$s=\beta_0+\beta_1 t+\beta_2 t^2$$

若测得如下数据:

t_i(秒)	1/30	2/30	3/30	4/30	5/30	6/30	7/30	8/30
s_i(厘米)	11.86	15.67	20.60	26.69	33.71	41.93	51.13	61.49
t_i	9/30	10/30	11/30	12/30	13/30	14/30	15/30	
s_i	72.90	85.44	99.08	113.77	129.54	146.48	165.06	

假定 s_i 服从正态 $N(\beta_0+\beta_1 t_i+\beta_2 t_i^2,\sigma)$，$(i=1,2,\cdots,15)$.

(1) 试确定 β_0、β_1、β_2 及 σ^2 的极大似然估计;

(2) 检验原假设 $H_0:\beta_2=0$;

(3) 给出 s 的置信水平 0.95 的预测区域.

4. 设测得周期为 2π 的函数 $y=f(x)$ 的数据如下:

x_i	15	30	45	60	75	90	105	120	135
y_i	1.31	1.84	2.33	2.41	2.24	2.59	2.12	2.38	2.98
x_i	150	165	180	195	210	225	240	255	270
y_i	3.44	3.51	3.33	2.89	2.01	0.02	−0.24	−1.23	−1.98
x_i	285	300	315	330	345	360			
y_i	−2.30	−2.21	−1.57	−1.63	−0.01	−0.62			

试确定下述表达式中系数的最小二乘法估计量：

$$\hat{y}=\beta_0+\beta_1\cos x+\beta_2\sin x+\beta_3\cos 2x+\beta_4\sin 2x.$$

5. 在例 9.5.2 的表 9.5.6 中，给出下列数据：

	y_{1k}	y_{2k}	y_{3k}	y_{4k}	y_{5k}	y_{6k}	y_{7k}	y_{8k}
$k=1$	795	756	891	858	956	860	806	736
$k=2$	805	750	885	850	965	870	811	730

试作方差分析并确定优的生产条件（即优的水平组合）。

6. 试证明(9.4.13)式及(9.4.15)式成立。

7. 试作变数变换，将§9.3之三的1—7各曲线方程化为线性方程。

8. 设 ξ_1 与 ξ_2 为独立的随机变数，数学期望分别为 θ 与 2θ，试求 θ 的最小二乘法估计与剩余平方和。

9. 设 $\mathbf{y}'=(y_1,y_2,y_3)$ 服从线性模型 $(\mathbf{y},\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I}_3)$，其中

$$E(y_i)=\beta_0+\beta_1x_i+\beta_2(3x_i^2-2),\qquad i=1,2,3.$$

$x_1=-1,x_2=0,x_3=1$. 试写出矩阵 $\mathbf{X}$，并求出 β_0,β_1,β_2 的最小二乘法估计，且证明当 $\beta_2=0$ 时，β_0 与 β_1 的最小二乘法估计量不变。

10. 设

$$y_i=\theta+\varepsilon_i,\qquad i=1,\cdots,m;$$
$$y_{m+i}=\theta+\phi+\varepsilon_{m+i},\qquad i=1,\cdots,m;$$
$$y_{2m+i}=\theta-2\phi+\varepsilon_{2m+i},\qquad i=1,\cdots,n.$$

假定 ε_i 之间互不相关，且有

$$E(\varepsilon_i)=0,D(\varepsilon_i)=\sigma^2,i=1,\cdots,2m+n.$$

试求 θ 及 ϕ 的最小二乘法估计. 试证当 $m=2n$ 时，$\hat{\theta}$ 与 $\hat{\phi}$ 互不相关。

11. 设 $\begin{cases} y_1 = a \quad\quad + e_1; \\ y_2 = 2a - b + e_2; \\ y_3 = a + 2b + e_3, \end{cases}$

其中 e_1, e_2, e_3 相互独立，且有

$$E(e_i) = 0, D(e_i) = \sigma^2, i = 1, 2, 3,$$

试求 a 及 b 的最小二乘法估计量.

12. 某医院用光电比色计检尿汞时，获得尿汞含量(mg/l)与消光系数读数的数据如下：

尿汞含量x_i	2	4	6	8	10
消光系数y_i	64	138	205	285	360

已知它们之间服从线性模型

$$E(y_i) = \beta_0 + \beta_1 x_i.$$

试求 β_0 和 β_1 的最小二乘法估计，并在显著性水平 $\alpha = 0.05$ 下，检验假设 $H_0: \beta_1 = 0$ 是否成立.

13. 今有四个物体，按下面方法称，得到如下数据：

x_1	x_2	x_3	x_4	y
1	1	1	1	20.2
1	−1	1	−1	8.0
1	1	−1	−1	9.2
1	−1	−1	1	1.4

其中 1 表示该物体放在天平左边，−1 表示放在天平的右边，y 是使天平达到平衡时，在右边所加砝码的重量，试估计这四个物体的重量 $\beta_i, i = 1, 2, 3, 4$.

第十章　随机过程引论

§10.1　随机过程的概念

一、随机过程的直观背景和定义

在前面几章讨论的随机现象，基本上可由一个或有穷多个随机变数来描述，它所考虑的基本事件可用一个或有穷多个数来表示. 在第一章的一开始，我们还指出，有些随机现象还必须研究它的发展变化过程. 因此，这一类随机现象仅只用一个或有穷多个随机变数去描述，就不能揭示其全部统计规律性，随机过程理论的诞生和发展，就是适应这一客观需要的.

下面举几个例子.

例 10.1.1　考虑某一电话站在正常工作条件下，某时刻 t 以前接到的呼叫次数 ξ，已知它是一个随机变数. 若我们考虑它随时间 t 变化的情形，则必须研究依赖于时间 t 的随机变数 ξ_t. 若以一天 24 小时计，则 t 从 0 变化到 24.

例 10.1.2　在数字通讯中，若传输过程是用数 0 和 1 两个码元通过编码来传递消息. 由于我们事先并不知道传送什么消息，因此，在某一时刻 t 它传送的是 0 还是 1，都不能事先预言，因而是一随机变数. 若我们进行长期观察，每隔 1 秒观察一次，则这个随机变数(用 ξ_n 表示)依赖于 n(秒). 其中 $n=0,1,2,\cdots$.

例 10.1.3　在地震勘探工作中，我们通过检波器把混有随机干扰的随时间变动的地层结构信号波记录下来. 如图 10.1.1 所示：在 O 点放炮，在 A 点记录仪把接收到的混有干扰的地震信号波记录下来，我们在相同条件下做了 n 次记录，则得到 n 个彼此有差

异的记录．在时间 t_0 观察它们的信号波的值 ξ_{t_0}，则可以发现它是不规则的，即 ξ_{t_0} 是一个随机变数．也就是说，混有随机干扰的地层结构信号波是一个依赖于时间 t 的随机变数 ξ_t．

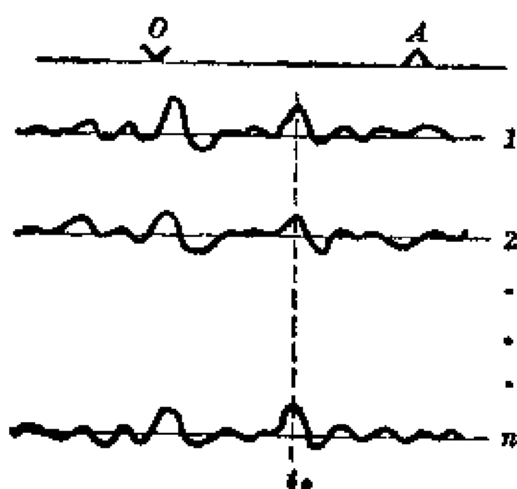


图 10.1.1

例 10.1.4 在例 10.1.3中，时间 t 是连续变化的（如在 0 至 5 秒内连续记录信号波）．但在对地震信号波进行数字处理时，则需按一定的时间间隔 Δ（如1毫秒）取样．这时，我们所研究的随机变数则依赖于 $n\Delta$（n 取非负整数）．固定 n_0 时，对应的 $\xi_{n_0\Delta}$ 为一随机变数．

总之，在实际问题中，不仅要求我们研究一个或有穷多个随机变数，而且要求我们研究一族无穷多个随机变数，我们将这样一族随机变数称为随机过程．

下面给出随机过程的正式定义：

定义 10.1.1 设给定概率空间 $(\Omega,\mathscr{F},P)$ 及指标集 T①，对每一 $t\in T$，有定义在 $(\Omega,\mathscr{F},P)$ 上的随机变数 $\xi_t(\omega)$，$\omega\in\Omega$，与之对应．则称依赖于 t 的一族随机变数 ξ_t 为随机过程，用符号 $\{\xi_t(\omega);t\in T\}$ 表示，如不产生混乱，简写为 $\{\xi_t\}$．其中指标 t 称为参数或时间，T 称为参数集．

这一章将研究能用随机过程描述的随机现象，叙述随机过程最初步的基本概念，至于需要更深入学习的读者，可参看[11]，[13]，[22]，[32]．

对于固定的时间 $t\,(t\in T)$，若随机变数 ξ_t 是离散型的，我们就

---

① 一般 T 为如下的指标集：$T=\{0,1,\cdots\}$，或 $T=\{\cdots,-1,0,1,\cdots\}$，或 $T=[a,b]$ 或 $T=R_1=(-\infty,\infty)$．

说随机过程$\{\xi_t\}$有一个离散状态空间，否则就说$\{\xi_t\}$有一个非离散状态空间. 参数集T可以是离散的(如$T=\{1,2,\cdots,n\}$或$T=\{1,2,\cdots\}$)或连续的 (如$T=\{t;t\geqslant0\}$或$T=R_1$). 若参数集T是离散的，则称随机过程是一个离散参数随机过程；若T是连续的，则称随机过程是一个连续参数随机过程.

因此，根据状态空间和参数集T的不同情况，可以把随机过程分成四类，列成下表：

表 10.1.1

随机过程 状态空间 / 参数集	离　散	非离散
离　散	离散参数链	离散参数随机过程
连　续	连续参数链	连续参数随机过程

把随机过程$\{\xi_t(\omega);\ t\in T\}$设想为一个作随机运动的质点$M$，以$\xi_t$表$M$在时刻$t$的位置，于是$\{\xi_t(\omega);\ t\in T\}$描述了$M$所作的随机运动的变化过程，这时，把状态空间的每一点x称作状态，$\xi_t=x$就可形象地说成“在时刻t质点M处于状态x”.

随机过程$\{\xi_t(\omega);\ t\in T\}$事实上是参数$t$和概率空间中的点$\omega$的二元函数. 对每一$t_0\in T$, $\xi_{t_0}(\omega)$是定义在$(\Omega,\mathscr{F},P)$上的随机变数. 对每一$\omega_0\in\Omega$, $\xi_t(\omega_0)$是定义在T上的一个普通的函数(实的或复的，也可以是实向量或复向量)，称为随机过程$\{\xi_t\}$对应于ω_0的一个样本函数，有时也称为轨道或现实. 图 10.1.1 画出了n个样本函数. 如果样本函数对几乎所有的ω(除了概率为 0 以外)为连续(或可微)的，则称样本函数几乎处处连续(或可微).

随机过程$\{\xi_t\}$的参数t可能不一定是我们通常所指的时间这一物理量，而可能是别的量，也可以是多维的量，例如考虑某一海面的波浪的浪高随时间的变化情况，则这时需用四维的参数来描

述(三维坐标 x,y,z, 及时间 t). 对于参数为多维的随机过程 $\{\xi_t\}$ 称为随机场.

二、随机过程的有穷维分布函数族

设 $\{\xi_t\}$ 是一随机过程,对于每一 $t\in T$, ξ_t 的分布函数为 $F_t(x)$, 即

$$F_t(x)=P\{\xi_t<x\}.$$

当任意给定 $t_1,t_2\in T$ 时,随机变数 ξ_{t_1} 和 ξ_{t_2} 的联合分布函数为 $F_{t_1,t_2}(x_1,x_2)$, 即

$$F_{t_1,t_2}(x_1,x_2)=P\{\xi_{t_1}<x_1,\xi_{t_2}<x_2\}.$$

一般地,当给定 $t_1,t_2,\cdots,t_n\in T$ 及 n 时, 随机变数 $\xi_{t_1},\xi_{t_2},\cdots,\xi_{t_n}$ 的 n 维联合分布函数为

$$F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n)=P\{\xi_{t_1}<x_1,\xi_{t_2}<x_2,\cdots,\xi_{t_n}<x_n\}.$$

一切上述分布函数的全体

$$\{F_{t_1,t_2,\cdots,t_n}(x_1,x_2,\cdots,x_n):t_1,\cdots,t_n\in T,n>0\}$$

称为 $\{\xi_t\}$ 的有穷维分布函数族, 它描述随机过程 $\{\xi_t\}$ 的统计规律.

读者不难验证,有穷维分布函数族满足下列条件:

(i) 对每一 $x_i(i=1,\cdots,n)$, $F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)$ 非负、不降且左连续.

(ii) $\lim\limits_{x_i\to-\infty}F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=0\quad(i=1,\cdots,n)$,

$$\lim_{\substack{x_1\to\infty\\ \cdots\cdots\\ x_n\to\infty}}F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=1.$$

(iii) 若 $x_i\leqslant y_i(i=1,\cdots,n)$, 则

$$F_{t_1,\cdots,t_n}(y_1,\cdots,y_n)-\sum_{i=1}^{n}F_{t_1,\cdots,t_n}(y_1,\cdots,y_{i-1},x_i,y_{i+1},\cdots,y_n)$$

$$+\sum_{\substack{i,j=1\\ i<j}}^{n}F_{t_1,\cdots,t_n}(y_1,\cdots,y_{i-1},x_i,y_{i+1},\cdots,y_{j-1},x_j,y_{j+1},\cdots,y_n)-\cdots$$

$$+(-1)^nF_{t_1,\cdots,t_n}(x_1,\cdots,x_n)\geqslant0$$

以及称为相容性条件的两个条件：

(iv) 对$(1,\cdots,n)$的任一排列$(\alpha_1,\cdots,\alpha_n)$，有

$$F_{t_1,\cdots,t_n}(x_1,\cdots,x_n)=F_{t_{\alpha_1},\cdots,t_{\alpha_n}}(x_{\alpha_1},\cdots,x_{\alpha_n}).$$

(v) 若$m<n$，则

$$F_{t_1,\cdots,t_m}(x_1,\cdots,x_m)=\lim_{x_{m+1},\cdots,x_n\to\infty}F_{t_1,\cdots,t_n}(x_1,\cdots,x_n).$$

上述命题的逆命题，即所谓随机过程的存在性定理，是一个非常重要且有名的定理，我们把它叙述如下：

设已给参数集T，对任意n及T的任意n个元素$t_1,\cdots,t_n$，有一n元函数$F_{t_1},\cdots,t_n(x_1,\cdots,x_n)$与之对应，并且满足上述(i)，(ii)，(iii)，(iv)，(v)五个条件（我们称这族函数为满足相容性条件的有穷维分布函数族），则必存在一随机过程$\{\xi_t\}$，使得它的有穷维分布函数族等于所给定的函数族.

对上述定理证明有兴趣的读者可参看[22]或[31].

要确定随机过程的分布函数族，在实际工作中往往是十分困难的，与第三章研究一维或多维随机变数时类似，我们可以转而研究随机过程的某种特征，它在许多实际问题和理论问题中都能很好地满足研究的目的.

(1) $\{\xi_t\}$的数学期望与方差

设随机过程$\{\xi_t\}$的一维分布函数为$F_t(x)$. 仿第三章，随机过程$\{\xi_t\}$的数学期望及方差分别定义为

$$E(\xi_t)=\int_{-\infty}^{\infty}x\,dF_t(x) \tag{10.1.1}$$

（当$\int_{-\infty}^{\infty}|x|\,dF_t(x)<\infty$时），及

$$D(\xi_t)=\int_{-\infty}^{\infty}[x-E(\xi_t)]^2dF_t(x) \tag{10.1.2}$$

$\left(当\int_{-\infty}^{\infty}[x-E(\xi_t)]^2dF_t(x)<\infty\right)$ 一般的说，$E(\xi_t)$与$D(\xi_t)$都是t的函数.

当$\{\xi_t\}$的状态空间是离散的时，则

$$E(\xi_t)=\sum_i x_i p_i(t), \qquad (10.1.3)$$

$$D(\xi_t)=\sum_i [x_i-E(\xi_t)]^2\, p_i(t), \qquad (10.1.4)$$

其中 $p_i(t)$ 为 ξ_t 的概率函数，即 $p_i(t)=P\{\xi_t=x_i\}$。

若对任一 $t\in T$，ξ_t 为连续型随机变数时，则

$$E(\xi_t)=\int_{-\infty}^{\infty} x f_t(x)\,dx, \qquad (10.1.5)$$

$$D(\xi_t)=\int_{-\infty}^{\infty} [x-E(\xi_t)]^2 f_t(x)\,dx, \qquad (10.1.6)$$

其中 $f_t(x)$ 为 ξ_t 的密度函数.

(2) $\{\xi_t\}$的相关函数和协方差函数

设随机过程$\{\xi_t\}$的二维分布函数为 $F_{t_1,t_2}(x_1,x_2)$，则$\{\xi_t\}$的相关函数定义为：对任意 $t_1, t_2\in T$

$$K(t_1,t_2)=E(\xi_{t_1}\cdot\xi_{t_2})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2\, dF_{t_1,t_2}(x_1,x_2) \qquad (10.1.7)$$

而把

$$\begin{aligned}R(t_1,t_2)&=E[\xi_{t_1}-E(\xi_{t_1})][\xi_{t_2}-E(\xi_{t_2})]\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x_1-E(\xi_{t_1})][x_2-E(\xi_{t_2})]\,dF_{t_1,t_2}(x_1,x_2)\end{aligned} \qquad (10.1.8)$$

称为$\{\xi_t\}$的协方差函数，有时亦称它为中心化相关函数. 自然，只有当(10.1.7)及(10.1.8)的积分绝对收敛时这两个式子才有定义.

相关函数与协方差函数之间有密切的关系：首先 $R(t_1,t_2)=K(t_1,t_2)-E(\xi_{t_1})E(\xi_{t_2})$ 其次，$\{\xi_t\}$的协方差函数是随机过程$\{\xi_t-E(\xi_t)\}$的相关函数，而$\{\xi_t\}$的相关函数可考虑作为关于随机过程$\{\xi_t e^{i\eta}\}$的协方差函数，其中 η 为均匀分布于$(-\pi,\pi)$的随机变数，并且与$\{\xi_t\}$独立. 上述结果，读者可自行验证.

相关函数 $K(t_1,t_2)$一般说是 t_1,t_2的二元函数. 可以证明 $K(t_1,t_2)$ 是非负定的. 即对任意 $n(n=1,2,\cdots,)$ 及数 $\alpha_i(i=1,\cdots,n)$

$$\sum_{i,\ j=1}^{n} K(t_i,t_j)\alpha_i\alpha_j \geqslant 0. \tag{10.1.9}$$

事实上,

$$\sum_{i,\ j=1}^{n} K(t_i,t_j)\alpha_i\alpha_j = E\left|\sum_{i=1}^{n}\xi_{t_i}\alpha_i\right|^2 \geqslant 0. \tag{10.1.10}$$

上述$\{\xi_t\}$的数学期望、方差、相关函数和协方差函数，分别描述随机过程在随时间 t 变化时的平均数、离散度和相关情况. 对于相关函数,我们还将在平稳过程部分着重讨论.

例 10.1.5 设 $g(t)$为如图10.1.2所示的以周期为L的矩形波,η 为服从两点分布的随机变数,其分布列如下:

η	1	-1
p_i	$\frac{1}{2}$	$\frac{1}{2}$

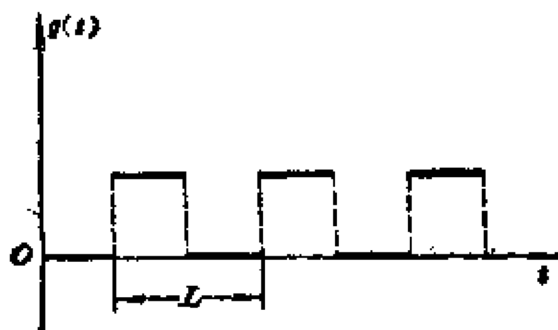


图 10.1.2

定义 ξ_t 如下:对任意 $t\in R_1$,

$$\xi_t=\eta g(t),$$

即$\{\xi_t\}$为具有随机振幅、周期为L的矩形波. 则有

$$E(\xi_t)=E[\eta g(t)]=g(t)\left[\frac{1}{2}+(-1)\cdot\frac{1}{2}\right]=0,$$

$$D(\xi_t)=E[\eta g(t)]^2=g^2(t)\cdot E(\eta^2)=g^2(t);$$

$$K(t_1,t_2)=E[\eta g(t_1)\cdot\eta g(t_2)]$$

$$=g(t_1)g(t_2)E(\eta^2)=g(t_1)\cdot g(t_2).$$

例 10.1.6 设 ξ,η 为相互独立、数学期望为 0、方差为 1 的随机变数，令

$$\xi_t=\xi+\eta t,$$

则

$$E(\xi_t)=E(\xi)+tE(\eta)=0;$$

$$D(\xi_t)=D(\xi+\eta t)=D(\xi)+t^2D(\eta)=1+t^2;$$

$$K(t_1,t_2)=E(\xi_{t_1}\xi_{t_2})$$

$$=E(\xi^2)+t_1\cdot t_2E(\eta^2)+(t_1+t_2)E(\xi)E(\eta)$$

$$=1+t_1t_2.$$

例 10.1.7① 随机过程 $\{\xi_t\}$ 定义如下：

考虑“随机点”发生在区间 $(0,t]$ 内的次数，若“随机点”在 $(0,t]$ 内发生的次数是偶数（把 0 也算作偶数），则令 $\xi_t=1$，若为奇数，则令 $\xi_t=-1$，如图 10.1.3 所示. 又设在时间区间 $(t_0,t_0+t]$ 内有 k 个“随机点”发生的概率与 t_0 无关并且为

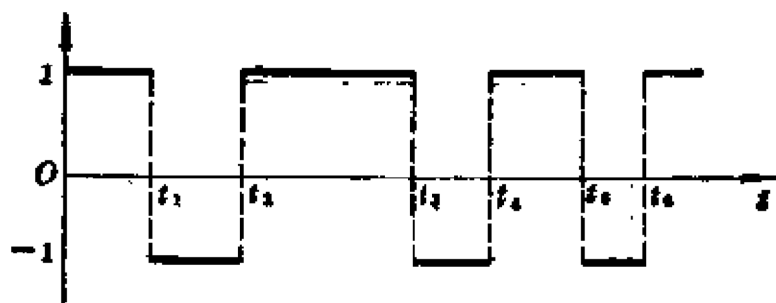


图 10.1.3

① 在此例中所述之“随机点”，在许多概率模型中，可以概括相当广泛的实际背景：电话总机接到的“呼叫”、数字编码通讯中出现的“误码”…，都可视为“随机点”. 在下一节我们还要进行讨论.

$$p_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},\quad (t\in R_1)$$

其中 $\lambda>0$, $k=0,1,2,\cdots$, 则有

$$P\{在(0,t]内发生偶数次“随机点”\}=p_0(t)+p_2(t)+\cdots$$
$$=e^{-\lambda t}\left\{1+\frac{(\lambda t)^2}{2!}+\frac{(\lambda t)^4}{4!}+\cdots\right\}=e^{-\lambda t}\cosh\lambda t$$
$$P\{在(0,t]内发生奇数次“随机点”\}=p_1(t)+p_3(t)+\cdots$$
$$=e^{-\lambda t}\left\{\lambda t+\frac{(\lambda t)^3}{3!}+\frac{(\lambda t)^5}{5!}+\cdots\right\}=e^{-\lambda t}\sinh\lambda t,$$

即

$$P\{\xi_t=1\}=e^{-\lambda t}\cosh\lambda t;$$
$$P\{\xi_t=-1\}=e^{-\lambda t}\sinh\lambda t.$$

故得

$$E(\xi_t)=1\cdot e^{-\lambda t}\cosh\lambda t+(-1)\cdot e^{-\lambda t}\sinh\lambda t$$
$$=e^{-\lambda t}(\cosh\lambda t-\sinh\lambda t)=e^{-\lambda t}\cdot e^{-\lambda t}$$
$$=e^{-2\lambda t}.$$

下面求$\{\xi_t\}$的相关函数和方差，首先求 ξ_{t_1}, ξ_{t_2} 的联合概率函数 $P\{\xi_{t_1}=x_1,\xi_{t_2}=x_2\}$, 其中 $t_1,t_2\in R_1$; $x_i=1$ 或 -1, $(i=1,2)$.

因为 $P\{\xi_{t_1}=x_1,\xi_{t_2}=x_2\}=P\{\xi_{t_1}=x_1\}\cdot P\{\xi_{t_2}=x_2/\xi_{t_1}=x_1\}$,因而只需求 $P\{\xi_{t_1}=x_1\}$ 及 $P\{\xi_{t_2}=x_2/\xi_{t_1}=x_1\}$ 便可.

设 $t_2>t_1$, 令 $t_2-t_1=\tau$, 则可得 $P\{\xi_{t_2}=1/\xi_{t_1}=1\}=P\{$在$(0,t_1]$内“随机点”发生偶数次的条件下，在$(0,t_2]$仍然发生偶数次$\}=P\{$在$(t_1,t_2]$内“随机点”发生偶数次$\}=P\{$在$(0,\tau]$内“随机点”发生偶数次$\}=e^{-\lambda\tau}\cosh\lambda\tau$,

而

$$P\{\xi_{t_1}=1\}=e^{-\lambda t_1}\cosh\lambda t_1,$$

故

$$P\{\xi_{t_1}=1,\xi_{t_2}=1\}=(e^{-\lambda t_1}\cosh\lambda t_1)(e^{-\lambda\tau}\cosh\lambda\tau).$$

同理可得

$P\{\xi_{t_1}=-1,\xi_{t_2}=-1\}=(e^{-\lambda t_1}\sinh\lambda t_1)(e^{-\lambda r}\cosh\lambda r)$;

$P\{\xi_{t_1}=-1,\xi_{t_2}=1\}=(e^{-\lambda t_1}\sinh\lambda t_1)(e^{-\lambda r}\sinh\lambda r)$

$P\{\xi_{t_1}=1,\xi_{t_2}=-1\}=(e^{-\lambda t_1}\cosh\lambda t_1)(e^{-\lambda r}\sinh\lambda r)$

这样一来，可求得

$$
\begin{aligned}
K(t_1,t_2)=E(\xi_{t_1}\xi_{t_2})
&=(e^{-\lambda t_1}\cosh\lambda t_1)(e^{-\lambda r}\cosh\lambda r)\\
&\quad+(-1)(-1)(e^{-\lambda t_1}\sinh\lambda t_1)(e^{-\lambda r}\cosh\lambda r)\\
&\quad+(-1)(e^{-\lambda t_1}\sinh\lambda t_1)(e^{-\lambda r}\sinh\lambda r)\\
&\quad+(-1)(e^{-\lambda t_1}\cosh\lambda t_1)(e^{-\lambda r}\sinh\lambda r)\\
&=e^{-\lambda(t_1+r)}(\cosh\lambda t_1\cosh\lambda r-\sinh\lambda t_1\sinh\lambda r\\
&\quad+\sinh\lambda t_1\cosh\lambda r-\cosh\lambda t_1\sinh\lambda r)\\
&=e^{-\lambda(t_1+r)}[\cosh\lambda(r-t_1)-\sinh\lambda(r-t_1)]\\
&=e^{-\lambda(t_1+r)}e^{-\lambda(r-t_1)}=e^{-2\lambda r}.
\end{aligned}
$$

即得

$$K(t_1,t_2)=e^{-2\lambda(t_2-t_1)}.$$

同理可得，当 $t_2<t_1$ 时，

$$K(t_1,t_2)=e^{2\lambda(t_2-t_1)}.$$

因而，对任意 $t_1,t_2\in R_1$，

$$K(t_1,t_2)=e^{-2\lambda|t_2-t_1|}.$$

由于

$$D(\xi_t)=K(t,t)-[E(\xi_t)]^2,$$

而 $K(t,t)=E(\xi_t^2)=1$，故 $\{\xi_t\}$ 的方差为

$$D(\xi_t)=1-[e^{-2\lambda t}]^2=1-e^{-4\lambda t}.$$

前面已说及，随机过程的数学期望与方差一般说来都是 t 的函数，而相关函数则是依赖于 t_1,t_2 的二元函数，但有一些随机过程的数学期望不依赖于 t 而为一常数，例如例10.1.5及例10.1.6;

其相关函数只依赖于 t_2-t_1, 如例 10.1.7. 即时间的平移不改变上述两个特征的数值. 具有这一特性的随机过程, 在研究飞机受湍流影响而产生随机振动、棉纱截面积的随机大小、数字通讯中的某些已编码信号等随机现象时起重要作用. 特别对于相关函数具有上述特性的随机过程, 我们还要着重讨论,详细情况见 § 10.4.

§ 10.2 几类重要的随机过程简介

由于理论研究和应用的不断深入与扩大, 各种类型的随机过程很多,讨论得也很深入,我们不可能在这本书中尽述, 这里仅只介绍在各种自然现象的侦测预报、无线电技术、自动控制和通讯等方面常用的一些随机过程.

一、独立增量过程(可加过程)

设 $\{\xi_t\}$ 为一随机过程,若对任意正整数 n 及 $t_i\in T, i=1,2,\cdots,n, t_1<t_2<\cdots<t_n$, 随机变数

$$\xi_{t_0}, \xi_{t_1}-\xi_{t_0}, \xi_{t_2}-\xi_{t_1}, \xi_{t_3}-\xi_{t_2}, \cdots, \xi_{t_n}-\xi_{t_{n-1}}$$

相互独立,则称 $\{\xi_t\}$ 为独立增量过程,又称为可加过程, 其中 t_0 为初始时刻,

例如: 某电话总机在时间区间 $[0, t]$ 内接到的呼叫次数设为 ξ_t, 并设 $P\{\xi_0=0\}=1$. 则 $\{\xi_t; t\in[0,\infty)\}$ 是一独立增量过程, 因为它在不相重叠的时间区间内发生呼叫的次数可以认为是相互独立的.

从定义我们看到,若 $\{\xi_t\}$ 是独立增量过程,则对任意 t, ξ_t 是相互独立随机变数之和.

二、正态随机过程(高斯过程)

设随机过程 $\{\xi_t\}$ 对任意正整数 n 及 $t_1, t_2, \cdots, t_n$, $t_i\in T(i=1,2,\cdots,n)$, 随机变数 $\xi_{t_1}, \xi_{t_2}, \cdots, \xi_{t_n}$ 的联合分布函数为 n 维正态分布,即有如下的密度函数:

$$f(x_1,x_2,\cdots,x_n)=\frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{a})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{a})} \qquad (10.2.1)$$

则称$\{\xi_t\}$为正态随机过程，其中

$$\mathbf{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix},\mathbf{a}=\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix},\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n}\\\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n}\\\cdots & \cdots & \cdots & \cdots\\\sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn}\end{pmatrix}.$$

$\boldsymbol{\Sigma}$为对称非负定矩阵，$\boldsymbol{\Sigma}^{-1}$表$\boldsymbol{\Sigma}$的逆矩阵，$|\boldsymbol{\Sigma}|$表$\boldsymbol{\Sigma}$的行列式，$(\quad)'$表向量或矩阵的转置.

特别当$n=2$，$\boldsymbol{\Sigma}=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$时，(10.2.1)式即为二维正态密度函数.

在无线电技术研究和应用中，真空管中的霰射效应(shot effect)，可以用正态随机过程来描述.

三、维纳过程

设随机过程$\{\xi_t(\omega);t\in[t_0,\infty)\}$满足下列条件：

(i) $\{\xi_t\}$是平稳独立增量，即$\{\xi_t\}$是独立增量过程，且对任意$t_1,t_2\in[t_0,\infty)$，$t_1<t_2$，及$h>0$，$\xi_{t_2+h}-\xi_{t_1+h}$有相同分布函数；

(ii) 对任意$t_1,t_2\in[t_0,\infty)$，$\xi_{t_1}-\xi_{t_2}$服从正态

$$N(0,\sqrt{|t_1-t_2|});$$

(iii) $P\{\xi_0=0\}=1$，

则称$\{\xi_t\}$为维纳过程.

维纳过程可以作为质点的布朗运动，以及电子电路中的理论噪声的一个很好的数学模型. 并且在通讯和控制理论中经常用到它.

维纳过程亦称为维纳-勒维过程或布朗运动过程.

四、泊松过程

设随机过程$\{\xi_t(\omega); t\in[t_0, \infty)\}$ $(t_0\geqslant 0)$的状态只取非负整数值，若满足下列三个条件：

(i) $P\{\xi_{t_0}=0\}=1$；

(ii) $\{\xi_t\}$为平稳独立增量过程；

(iii) 对任意 $t_1, t_2\in[t_0, \infty)$, $t_1<t_2$，对应的随机变数的增量 $\xi_{t_2}-\xi_{t_1}$ 服从数学期望为 $\lambda(t_2-t_1)$的泊松分布，即对于 $k=0, 1, \cdots$有

$$P\{\xi_{t_2}-\xi_{t_1}=k\}=e^{-\lambda(t_2-t_1)}\frac{[\lambda(t_2-t_1)]^k}{k!};$$

$$E(\xi_{t_2}-\xi_{t_1})=\lambda(t_2-t_1); D(\xi_{t_2}-\xi_{t_1})=\lambda(t_2-t_1).$$

则称$\{\xi_t\}$为泊松过程.

对于公用事业、通讯等许多实际问题，都可以用泊松过程来描述，因为这些实际问题都可满足下面将要叙述的定理 10.2.1 的四个条件. 例如某电话总机在时间区间$[0, t]$内接到的呼叫次数、某服务窗口排队的人数、数字通讯中已编码信号的误码个数等，都能很好的满足所说四个条件，因而都可看成是泊松过程.

五、随机点过程与计数过程

在例 10.1.7 中，我们提到了"随机点"这一概念，这里我们稍作介绍.

在客观世界中，存在着这样一类随机现象，它们发生的时间、地点或相联系的某些属性，常可归结为某一空间 $\mathscr{X}$ 中的点的随机发生，这种点的发生就构成所谓随机点过程.

例 10.2.1 一天中某电话总机接到的呼叫形成一个随机点过程. $\mathscr{X}$ 为时间区间$[0, 24]$，每一次呼叫发生的时间就是一个随机点，这个点过程的一个现实(样本函数)是一时间的序列$\{t_1, t_2, \cdots, t_N\}$，其中$0\leqslant t_1<t_2\cdots<t_N\leqslant 24$，$t_n$ 表示第 n 次呼叫发生的时

间，N 则表示一天内呼叫发生的次数.

例 10.2.2 某一台机器在时刻 t_0 开始工作，我们在时间区间 $[t_0, t_0+T]$ 内考虑机器发生故障而停止工作这一事件. 假设当机器发生故障后立即修理或换上新的同样的零件后继续工作，则机器在时间区间 $[t_0, t_0+T]$ 内发生故障而停止工作这事件形成一个随机点过程.

例 10.2.3 设从时间 t_0 开始的长为 T 的时间区间内考虑某地区发生地震这一事件(在一定的震级范围内，每一次地震的震中、发震时间)，它可以根据各台、站的观察结果记录下来. 由于震中的广度最多几公里，而震动的持续时间一般以秒计，这相对于我们研究的整个地区和时间范围来说是很小的，所以可以理想化为一点. 这样，每一地震发生这事件，可用一个三维向量 (x, y, t) 来表示，其中 x, y 表示震中的经度和纬度，t 表时间. 描述地震发生的事件序列 $(x_1, y_1, t_1), (x_2, y_2, t_2), \cdots (x_N, y_N, t_N)$ 构成一多维随机点过程.

由上所述，随机点过程是一种类型的随机过程，人们关心的首先是代表随机事件发生的时间、地点或状态的点的统计规律，这些统计规律包括如下两个方面，

(1) 计数性质：考虑这种随机的点落在 $\mathscr{X}$ 的某一部分的个数，当 $\mathscr{X}$ 是时间轴 $[0, \infty)$ 时，令 N_t 表示在 $[0, t)$ 内发生这种点的数目，即随机事件发生的次数，则 $\{N_t; t\in[0, \infty)\}$ 是一个取非负整数值的递增的随机过程，一般称为伴随这随机点过程的计数过程(counting process).

(2) 在一维的情形($\mathscr{X}=R_1$)，考虑一随机点的发生与相继另一随机点发生的间距，令 T_1 表示初始时刻到第一个随机点发生的时间间隔，T_n 表第 $n-1$ 个随机点到第 n 个随机点发生的时间间隔，则 $T_1, T_2, \cdots$ 是一随机序列，$T_n (n=1, 2, \cdots)$ 有时称为点间间距，

当 T_2, $T_3\cdots$同分布时(可能 T_1 不同), 重要的问题是考虑 $T_n(n=2,3,\cdots)$的分布,即考虑点间间距分布.

随机点过程在早期称为随机事件流, 它早期的研究出现于排队论和运筹学等分支中, 近年来对随机点过程的基础理论研究和应用正蓬勃开展,这方面的内容可参看[34].

关于计数过程有如下定理,

定理 10.2.1 若计数过程$\{N_t;t\in[t_0,\infty)\}$满足下列条件:

(1) $P\{N_{t_0}=0\}=1$, 即以概率为 1 地在初始时刻处于状态为 0;

(2) $\{N_t\}$为平稳独立增量过程;

(3) 对任意 $t>t_0$,$0<P\{N_t>0\}<1$. 换句话说, 在时间区间$[t_0,t)$内有"随机点"发生的概率为正, 但不是一定有"随机点"发生;

(4) 对任意 $t\geqslant t_0$

$$\lim_{h\to 0}\frac{P\{N_{t+h}-N_t\geqslant 2\}}{P\{N_{t+h}-N_t=1\}}=0,$$

换句话说,在充分小的时间区间内,至多有一个"随机点"发生. 则$\{N_t\}$为泊松过程.

定理 10.2.1 的证明可参看定理 2.1.2 的证明.

随机过程还有许多类型,不能一一加以介绍,但对于应用中比较有用的马氏链与平稳过程,我们将在下两节加以介绍.

§10.3 马 氏 过 程

我们在这一节着重介绍应用比较广泛且较基本的马氏链.

一、马氏链的定义及例子

定义 10.3.1 设随机序列$\{\xi_n;n\in N\}$满足下列两个条件(其中N一般包含有穷或可列无穷个非负整数):

(i) 对每一 $n\in N$，ξ_n 只取非负整数值；

(ii) 对任意非负整数 m，l，k 及任意非负整数 $s_1, s_2, \cdots, s_l$ $(s_1<s_2<\cdots<s_l)$ 及 $i_{s_1}, i_{s_2}, \cdots, i_{s_l}, i_m, i_{m+k}$ $(m>s_l)$，有

$$P\{\xi_{m+k}=i_{m+k}/\xi_{s_1}=i_{s_1}, \xi_{s_2}=i_{s_2}\cdots, \xi_{s_l}=i_{s_l}, \xi_m=i_m\}$$
$$=P\{\xi_{m+k}=i_{m+k}/\xi_m=i_m\}. \tag{10.3.1}$$

当(10.3.1)式左方有意义时．则称$\{\xi_n\}$为马尔科夫链，或简称为马氏链.

在下面讨论马氏链的时候，我们总假设构成条件的事件有正的概率，例如事件$\{\xi_m=i_m\}$应满足 $P\{\xi_m=i_m\}>0$.

马氏链是较早出现的马氏过程中的一类．它早期是用下述比较直观、但一直沿用到现在的语言叙述的.

考虑一系列随机试验，每一次试验只有有穷个或可列无穷个基本事件 $A_i(i=1,2,\cdots)$ 可能发生，且每次试验将发生而且仅发生这些事件中的一个事件.若事件 A_i 发生，就说系统处于状态 A_i（或简称为状态 i）．为方便起见，以 $A_i^{(k)}$ 表示第 k 次试验结果处于状态 A_i 这一事件．我们还常常用"时刻"这一用语，设想每隔一个单位时间进行一次试验，这样，$A_i^{(k)}$就可说成"于时刻 k 系统处于状态 i"这一事件，而且把 $P\{A_{i+s}^{(k+l)}/A_i^{(k)}\}$ 称为系统在时刻 k 处于状态 i，经过 l 步后系统转移到状态 $i+s$ 的转移概率．这样一来，(10.3.1)式右边就可称为系统在时刻m处于状态 i_m，经 k 步后系统处于 i_{m+k} 的转移概率．(10.3.1)就可直观地叙述为：已知系统现在(时刻 m)处于状态 i_m，那么，将来(时刻 $m+k$)的状态与过去(时刻 $s_1, s_2, \cdots, s_l$)的状态无关.或简述为：已知系统的"现在"，那么，"将来"与"过去"无关．这种特性我们称为马氏性[①].

①：我们不能把马氏性解释为"将来"决定于"现在"而与"过去"无关，它只说在"知道现在"的条件下，过去的信息对推断将来的概率性质不起作用.

转移概率 $P\{\xi_{m+k}=j/\xi_m=i\}$ 记为 $p_{ij}(m,k)$①. 即

$$p_{ij}(m,k)=P\{\xi_{m+k}=j/\xi_m=i\}\,(i,j=0,1,2,\cdots).$$

我们把上述转移概率写成矩阵形式，并用 $\mathbf{P}(m,k)$ 表示，即

$$\mathbf{P}(m,k)=(p_{ij}(m,k))$$

$$=\begin{pmatrix} p_{00}(m,k) & p_{01}(m,k) & p_{02}(m,k) & \cdots & p_{0n}(m,k)\cdots \\ p_{10}(m,k) & p_{11}(m,k) & p_{12}(m,k) & \cdots & p_{1n}(m,k)\cdots \\ p_{20}(m,k) & p_{21}(m,k) & p_{22}(m,k) & \cdots & p_{2n}(m,k)\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ p_{n0}(m,k) & p_{n1}(m,k) & p_{n2}(m,k) & \cdots & p_{nn}(m,k)\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \end{pmatrix}$$

并称 $\mathbf{P}(m,k)$ 为 $\{\xi_n\}$ 的 k **步转移概率矩阵**，特别称 $\mathbf{P}(m,1)$ 为**转移概率矩阵**.

由 $p_{ij}(m,k)$ 的定义，显然

$$\left.\begin{aligned} & p_{ij}(m,k)\geqslant 0, \\ & \sum_j p_{ij}(m,k)=1,\ i=0,1,2,\cdots \end{aligned}\right\} \tag{10.3.2}$$

定理 10.3.1 设 $\{\xi_n\}$ 为一马氏链，则对任意非负整数 k, m, l, 有

$$p_{ij}(m,k+l)=\sum_r p_{ir}(m,k)\,p_{rj}(m+k,l). \tag{10.3.3}$$

或采用矩阵写法，则有

$$\mathbf{P}(m,k+l)=\mathbf{P}(m,k)\mathbf{P}(m+k,l).$$

(10.3.3)式称为切普曼-柯尔莫哥洛夫方程.

---

① 对于初学者要特别注意，不要忘记转移概率原意是条件概率，这里用了它的形象说法.

在前面几章，我们用 $P(k/i)$ 表示条件概率，在这一章里，条件概率沿用大多数著作的符号.

证 $p_{ij}(m,k+l)=P\{\xi_{m+k+l}=j/\xi_m=i\}$

$$=\frac{P\{\xi_m=i,\xi_{m+k+l}=j\}}{P\{\xi_m=i\}}$$

$$=\sum_r\frac{P\{\xi_m=i,\xi_{m+k}=r,\xi_{m+k+l}=j\}}{P\{\xi_m=i,\xi_{m+k}=r\}}\cdot\frac{P\{\xi_m=i,\xi_{m+k}=r\}}{P\{\xi_m=i\}}$$

$$=\sum_r P\{\xi_{m+k+l}=j/\xi_m=i,\xi_{m+k}=r\}\cdot P\{\xi_{m+k}=r/\xi_m=i\}$$

利用马氏性得 $P\{\xi_{m+k+l}=j/\xi_m=i,\xi_{m+k}=r\}=p_{rj}(m+k,l)$，而 $P\{\xi_{m+k}=r/\xi_m=i\}=p_{ir}(m,k)$，故得

$$p_{ij}(m,k+l)=\sum_r p_{ir}(m,k)p_{rj}(m+k,l).\ ∎$$

用归纳法并应用切普曼-柯尔莫哥洛夫方程可证得下面公式成立：

$$\left.\begin{aligned}p_{ij}(m,k+1)&=\sum_{s_1,s_2,\cdots,s_k}p_{is_1}(m,1)p_{s_1s_2}(m+1,1)\cdots p_{s_kj}(m+k,1);\\ p_{ij}(m,k+1)&=\sum_r p_{ir}(m,1)p_{rj}(m+1,k)\\ &=\sum_r p_{ir}(m,k)p_{rj}(m+k,1).\end{aligned}\right\}\quad(10.3.4)$$

对每一非负整数 k，记

$$p_j(k)=P\{\xi_k=j\},\ (j=0,1,2,\cdots);$$

$$p_j=p_j(0).$$

称 $p_j(k)$ 为绝对概率，特别称 p_j 为初始概率. 由全概率公式可得

$$\begin{aligned}p_j(k+1)&=\sum_i p_ip_{ij}(0,k+1)\\ &=\sum_i p_i(k)p_{ij}(k,1).\end{aligned}\quad(10.3.5)$$

(10.3.5)式的前一个等式说明：绝对概率 $p_j(k)$ 可由初始概率

和转移概率所决定.

定理 10.3.2 设$\{\xi_n\}$为一马氏链，则对于任意非负整数$s_1<s_2<\cdots<s_k$,及非负整数k,有

$$P\{\xi_{s_1}=i_{s_1},\xi_{s_2}=i_{s_2},\cdots,\xi_{s_k}=i_{s_k}\}$$
$$=p_{i_{s_1}}(s_1)\cdot p_{i_{s_1}i_{s_2}}(s_1,s_2-s_1)\cdots p_{i_{s_{k-1}}i_{s_k}}(s_{k-1},s_k-s_{k-1}). \tag{10.3.6}$$

证 由马氏性及概率的乘法公式可得

$$P\{\xi_{s_1}=i_{s_1},\xi_{s_2}=i_{s_2},\cdots,\xi_{s_k}=i_{s_k}\}$$
$$=P\{\xi_{s_1}=i_{s_1}\}\cdot P\{\xi_{s_2}=i_{s_2}/\xi_{s_1}=i_{s_1}\}\cdots$$
$$\cdot P\{\xi_{s_k}=i_{s_k}/\xi_{s_1}=i_{s_1},\xi_{s_2}=i_{s_2},\cdots,\xi_{s_{k-1}}=i_{s_{k-1}}\}$$
$$=P\{\xi_{s_1}=i_{s_1}\}\cdot P\{\xi_{s_2}=i_{s_2}/\xi_{s_1}=i_{s_1}\}\cdots$$
$$\cdot P\{\xi_{s_k}=i_{s_k}/\xi_{s_{k-1}}=i_{s_{k-1}}\}$$
$$=p_{i_{s_1}}(s_1)p_{i_{s_1}i_{s_2}}(s_1,s_2-s_1)\cdots\cdots p_{i_{s_{k-1}}i_{s_k}}(s_{k-1},s_k-s_{k-1}).$$ ∎

从(10.3.5)及(10.3.6)两式可见，对于马氏链，任意有限维概率函数完全由初始概率及转移概率所决定，也就是说，对于马氏链说来，初始概率及转移概率(只牵涉到二维概率函数)就可以完备的描述马氏链的统计特性.

二、齐次马氏链

定义 10.3.2 称马氏链$\{\xi_n\}$为齐次的，如果它的转移概率矩阵$\mathbf{P}(m,1)$与m无关，记为$\mathbf{P}(1)$. 即对任意非负整数m，有

$$p_{ij}(m,1)=p_{ij}(1)(i,j=0,1,2,\cdots).$$

由(10.3.4)式的第一个等式可得：

(1) 马氏链$\{\xi_n\}$是齐次的充分必要条件为：任意k步转移概率$p_{ij}(m,k)$都与m无关. 此时，记k步转移概率矩阵为$\mathbf{P}(k)$;

(2) 若$\{\xi_n\}$为齐次马氏链，则任意k步转移概率$p_{ij}(k)$，皆可由一步转移概率决定. 因此，$p_{ij}(1)$有特殊意义.

对齐次马氏链，我们记

$$p_{ij}=p_{ij}(1),$$

$$p_{ij}(k)=P\{\xi_{m+k}=j/\xi_m=i\}.$$

这样一来，对于齐次马氏链来说，只须将 $p_{ij}(m,k)$ 改为 $p_{ij}(k)$，前面所叙述的诸命题仍然成立，我们特别写出下列诸式：

$$\left.\begin{array}{l} p_{ij}\geqslant 0 \qquad (i,j=0,1,2,\cdots), \\ \sum_j p_{ij}=1, \text{对任意 } i=0,1,\cdots\text{皆成立} \end{array}\right\}; \tag{10.3.7}$$

$$p_{ij}(k+1)=\sum_{s_1,s_2,\cdots,s_k} p_{is_1}p_{s_1s_2}\cdots p_{s_kj}; \tag{10.3.8}$$

$$p_{ij}(k+l)=\sum_r p_{ir}(k)p_{rj}(l); \tag{10.3.9}$$

$$p_j(k+1)=\sum_i p_i p_{ij}(k+1)=\sum_i p_i(k)p_{ij}. \tag{10.3.10}$$

在(10.3.8)式中，特别当 $k=1$ 时，有

$$p_{ij}(2)=\sum_r p_{ir}p_{rj}.$$

把它写成矩阵的形式为

$$\mathbf{P}(2)=[\mathbf{P}(1)]^2.$$

用数学归纳法可得

$$\mathbf{P}(n)=[\mathbf{P}(1)]^n. \tag{10.3.11}$$

例 10.3.1 考虑贝努里试验，每次试验有两种状态：$A_1=A$，$A_2=\bar{A}$. 由于试验是独立的，且 $P(A_1)=p$，$p(A_2)=q$，$q=1-p$. 因此，在第 k 次试验出现 A_i（用 $\xi_k=i$ 表示）的条件下第 $k+1$ 次试验出现 A_j 的条件概率与 k 无关且等于 $P(A_j)$，这说明贝努里试验构成一个齐次马氏链，并且有

$$p_{11}=p_{21}=p;$$

$$p_{12}=p_{22}=q.$$

其转移概率矩阵为

$$\mathbf{P}(1)=\begin{pmatrix} p & q \\ p & q \end{pmatrix}.$$

并可算得

$$\mathbf{P}(2)=\begin{pmatrix} p & q \\ p & q \end{pmatrix}=\mathbf{P}(1).$$

更一般地，有

$$\mathbf{P}(n)=\begin{pmatrix} p & q \\ p & q \end{pmatrix}=\mathbf{P}(1) \qquad (n\geqslant 2).$$

例 10.3.2 考虑所谓"一个分子在两个附着壁之间的随机游动"这一模型．这是物理学中常常出现的一种现象：一个分子能处在数轴上 $1,2,\cdots,s$ 这 s 点中任一位置．一旦这分子到达 1 这一点处，它就一直留在那里不动，对 s 点也有这种情况．将 1 及 s 这两点称为附着壁．如果在时刻 t 的时候这分子处在点 i 处，其中 $2\leqslant i\leqslant s-1$，那么，它在一个单位时间后转到 $i+1$ 处(称为向前)的概率为 $p(0<p<1)$，转到点 $i-1$ 处(称为向后)的概率为 $q,q=1-p$．我们规定这分子只能"向前"或"向后"游动一步，并且经过一单位时间，它必须"向前"或"向后"作游动(1 和 s 处除外)．这样一个随机游动构成一个马氏链，并且是齐次的，其转移概率为：

$$p_{11}=p_{ss}=1.$$

且对于 $i\,(2\leqslant i\leqslant s-1)$ 有

$$p_{ir}=\begin{cases} p & \text{当 } r=i+1 \\ q & \text{当 } r=i-1. \\ 0 & \text{其他} \end{cases}$$

即转移概率矩阵为如下形式：

$$\mathbf{P}(1)=\begin{pmatrix}1 & 0 & 0 & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 0\\ q & 0 & p & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 0\\ 0 & q & 0 & p & 0 & \cdots & 0 & \cdots & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & q & 0 & p\\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 1\end{pmatrix}$$

让我们特别考虑当 $s=3$ 的情形，这时有

$$\mathbf{P}(1)=\begin{pmatrix}1 & 0 & 0\\ q & 0 & p\\ 0 & 0 & 1\end{pmatrix}.$$

可得当 $n\geqslant 2$ 时，

$$\mathbf{P}(n)=\mathbf{P}(1).$$

例 10.3.3 考虑没有附着壁的只有三个状态的随机游动：

类似于例 10.3.2 中的分子随机游动，但不假设有两个附着壁，即设 $p_{11}=q$，$p_{33}=p$. 其余假设与例 10.3.2 相同. 这时，转移概率矩阵为

$$\mathbf{P}(1)=\begin{pmatrix}q & p & 0\\ q & 0 & p\\ 0 & q & p\end{pmatrix}.$$

可求得

$$\mathbf{P}(2)=[\mathbf{P}(1)]^2=\begin{pmatrix}q^2+pq & pq & p^2\\ q^2 & 2pq & p^2\\ q^2 & pq & p^2+pq\end{pmatrix}$$

$$=\begin{pmatrix}q & pq & p^2\\ q^2 & 2pq & p^2\\ q^2 & pq & p\end{pmatrix}.$$

三、遍历性与平稳分布

定义 10.3.3 设$\{\xi_n\}$为齐次马氏链，若对一切状态 i 及 j，存在不依赖于 i 的常数 π_j，使得

$$\lim_{n\to\infty} p_{ij}(n)=\pi_j \tag{10.3.12}$$

则称马氏链$\{\xi_n\}$具有遍历性.

遍历性可解释为：不论系统自那一状态出发，当转移的"步长"n 充分大时，转移到状态 j 的概率近似于某个常数 π_j. 因此可用 π_j 近似 $p_{ij}(n)$，只要 n 充分大即可.

定理 10.3.3 对有限($s+1$个)状态的齐次马氏链$\{\xi_n\}$，若存在正整数 n_0，使对一切 $i,j=0,1,2,\cdots,s$，有

$$p_{ij}(n_0)>0,$$

则此马氏链$\{\xi_n\}$是遍历的，而且(10.3.12)式中之 $\pi_j(j=0,1,2,\cdots,s)$是如下方程组

$$\pi_j=\sum_{i=0}^{s}\pi_i p_{ij} \qquad (j=0,1,2,\cdots,s) \tag{10.3.13}$$

在满足条件：

$$\pi_j>0 \qquad (j=0,1,2,\cdots,s);$$

$$\sum_{j=0}^{s}\pi_j=1$$

下的唯一解.

定理 10.3.3 的证明可参看[1]，[5].

定义 10.3.4 设$\{\xi_n\}$为有限($s+1$个)状态的齐次马氏链，若初始概率 p_j 满足下式：

$$p_j=\sum_{i=0}^{s}p_i p_{ij} \qquad (j=0,1,2,\cdots,s), \tag{10.3.14}$$

则称$\{\xi_n\}$为平稳的，$p_j(j=0,1,2,\cdots,s)$称为$\{\xi_n\}$的一个平稳分布.

由(10.3.10)式，令 $k=0$ 得

$$p_j(1)=\sum_{i=0}^{s}p_ip_{ij},$$

又由(10.3.14)式及上式得，对任意 $j=0,1,2,\cdots,s$

$$p_j(1)=p_j.$$

同理可得，当 $k=1$ 时，由(10.3.10)式

$$p_j(2)=\sum_{i}^{s}p_i(1)p_{ij},$$

又由(10.3.14)式并注意到 $p_j(1)=p_j$，故有

$$p_j(2)=p_j(1)=p_j.$$

用同样方法可证得，对任意 $j=0,1,2,\cdots,s$，

$$p_j=p_j(1)=p_j(2)=\cdots \tag{10.3.15}$$

由(10.3.15)式看到，若 $\{\xi_n\}$ 是平稳的，则其初始概率 p_j 等于绝对概率 $p_j(n)$ $(n=1,2,\cdots)$. 换句话说，平稳齐次马氏链在任一时刻处于状态为 j 的概率都相等. 这就是"平稳"的含义. 读者可以证明，(10.3.14)式与(10.3.15)式是等价的.

例 10.3.4 如例 10.3.3，我们可得，对任意 $i,j=1,2,3$，找到 $n_0=2$，使得

$$p_{ij}(2)>0.$$

由定理 10.3.3 可知，例 10.3.3 所述的随机游动构成的马氏链是遍历的. 即对任意 $i=1,2,3$，

$$\lim_{n\to\infty}p_{ij}(n)=\pi_j \quad (j=1,2,3).$$

且其中 π_j 可根据(10.3.13)式求得如下：由

$$\pi_1=\pi_1q+\pi_2q,$$

$$\pi_2=\pi_1p+\pi_3q,$$

$$\pi_3=(\pi_2+\pi_3)p,$$

及条件

$$\pi_j>0 \quad (j=1,2,3);$$
$$\pi_1+\pi_2+\pi_3=1.$$

可得

$$\pi_2=\frac{p}{q}\pi_1;$$
$$\pi_3=\left(\frac{p}{q}\right)^2\pi_1;$$
$$\pi_1\left[1+\frac{p}{q}+\left(\frac{p}{q}\right)^2\right]=1.$$

当$p=q=\frac{1}{2}$时可得，

$$\pi_1=\pi_2=\pi_3=\frac{1}{3}.$$

这表示在极限情况，处于状态1或2或3皆是等可能的. 这个结果很直观，因为此时分子在任一状态“向前”或“向后”的可能性相等$\left(p=q=\frac{1}{2}\right)$，因而当这分子游动相当长的时间后处于任一状态的可能性也是相同的.

当$p\neq q$时，可得

$$\pi_j=\frac{1-\frac{p}{q}}{1-\left(\frac{p}{q}\right)^3}\left(\frac{p}{q}\right)^{j-1} \qquad (j=1,2,3).$$

从上式看到，当$p>q$时，π_j随j的增加而增大；当$p<q$时则相反. 例如，当$\frac{p}{q}=2$时，

$$\pi_1=\frac{1}{7},\ \pi_2=\frac{2}{7},\ \pi_3=\frac{4}{7}.$$

而在$\frac{p}{q}=\frac{1}{2}$时，

$$\pi_1=\frac{4}{7},\ \pi_2=\frac{2}{7},\ \pi_3=\frac{1}{7}.$$

上面的结果也是很直观的，当 $p>q$ 时，表示这分子"向前"的可能性大于"向后"的可能性，因而当这分子游动相当长时间后，处于下一个状态(例如状态 3)的可能性比处于前一个状态(例如状态 2)的可能性大. 当 $p<q$ 时，则有相反的结果.

上面我们叙述了关于马氏链的遍历性和平稳性的概念，定理 10.3.3 给出了一个有限齐次马氏链具有遍历性的充分条件. 我们还要指出，当有限齐次马氏链 $\{\xi_n\}$ 满足定理 10.3.3 的条件，并且其初始概率 p_j 等于由(10.3.12)式确定的 π_j 时，则可证明 $\{\xi_n\}$ 为平稳的. 这结果读者不难从上面的叙述过程中得到.

具有马氏性的更一般的过程 $\{\xi_t;\ t\in T\}$ (其中 T 为有穷或无穷区间)称为马尔科夫过程，或简称为马氏过程. 即，若随机过程 $\{\xi_t\}$ 满足如下条件：对任意的正整数 n 及 $t_i\in T\ (i=1, 2, \cdots, n)$，$t_1<t_2\cdots<t_n$，以及任意实数 $x_1, x_2, \cdots, x_{n-2}, x, y$，如下等式成立

$$P\{\xi_{t_n}<y/\xi_{t_{n-1}}=x, \xi_{t_{n-2}}=x_{n-2}, \cdots, \xi_{t_1}=x_1\}$$
$$=P\{\xi_{t_n}<y/\xi_{t_{n-1}}=x\},$$

则称 $\{\xi_t\}$ 为马氏过程.

对于马氏过程的讨论，有兴趣的读者可参阅[22].

§ 10.4 平稳随机过程

一、平稳随机过程的定义及例

定义 10.4.1 设 $\{\xi_t(\omega): t\in R_1\}$ 为一随机过程，若对任意正整数 n 及实数 $r, t_1, t_2, \cdots, t_n$，随机变数 $\xi_{t_1}, \xi_{t_2}, \cdots, \xi_{t_n}$ 的 n 维分布函数与 $\xi_{t_1+r}, \xi_{t_2+r}, \cdots, \xi_{t_n+r}$ 的 n 维分布函数相同，即

$$F_{t_1,t_2,\cdots,t_n}(x_1, x_2, \cdots, x_n)=F_{t_1+r,t_2+r,\cdots,t_n+r}(x_1, x_2, \cdots, x_n), \tag{10.4.1}$$

则称$\{\xi_t\}$为强平稳过程. 换句话说，若随机过程$\{\xi_t\}$的任意有限维分布函数沿时间轴作平移时不改变. 则称$\{\xi_t\}$为强平稳过程.

定义 10.4.2 若随机过程$\{\xi_t(\omega);t\in R_1\}$满足如下两个条件:

$$\left.\begin{array}{ll}\text{(i)} & \text{对任意 } t\in R_1, E(\xi_t)^2<\infty; \\ \text{(ii)} & K(t_1,t_2)=B(t_1-t_2)^{①}.\end{array}\right\} \tag{10.4.2}$$

则称$\{\xi_t\}$为弱平稳过程.

$K(t_1,t_2)=B(t_1-t_2)$等价于$K(t_1+\tau,\ t_2+\tau)=K(t_1,\ t_2)$，或者$K(t+\tau,t)=B(\tau)$，其中$\tau\in R_1$.

显然，若$\{\xi_t\}$为强平稳过程，且$E|\xi_t|^2<\infty$，则$\{\xi_t\}$必为弱平稳过程，事实上，对任意t_1,t_2及$\tau\in R_1$,

$$\begin{aligned}K(t_1+\tau,t_2+\tau)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2dF_{t_1+\tau,t_2+\tau}(x_1,x_2)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1x_2dF_{t_1,t_2}(x_1,x_2)\\&=K(t_1,t_2).\end{aligned}$$

上述命题的逆命题则不一定成立. 但是，当$\{\xi_t\}$为零均值的正态过程时，则强平稳过程必为弱平稳过程，反之亦然. 这从(10.2.1)式可以得到.

在许多问题中，我们还常见到所谓平稳(强或弱)增量过程，它的定义如下:

若对每一$t,h\in R_1$,随机过程$\{\xi_t\}$的增量$\{\xi_{t+h}-\xi_t\}$为平稳(强或弱)过程，则称$\{\xi_t\}$为平稳(强或弱)增量过程.

例 10.4.1 考虑一个具有随机相位的余弦波，它由如下定义的随机过程描述,

$$\xi_t=\cos(\lambda t+\theta)\qquad(t\in R_1),$$

① 当$K(t_1,t_2)$只依赖于t_1-t_2时，我们把它记为$B(t_1-t_2)$. 这时，$K(t+\tau,t)=B(\tau)$.

其中 λ 是常数，θ 为一个具有有限二阶矩的随机变数．则可求得其数学期望和相关函数如下：

$$
\begin{aligned}
E(\xi_t) &= E[\cos(\lambda t+\theta)] \\
&= E(\cos\lambda t\cos\theta - \sin\lambda t\sin\theta) \\
&= \cos\lambda t E(\cos\theta) - \sin\lambda t E(\sin\theta);
\end{aligned}
$$

$$
\begin{aligned}
K(t_1, t_2) &= E[\cos(\lambda t_1+\theta)\cos(\lambda t_2+\theta)] \\
&= E\left[\frac{1}{2}\cos(\lambda t_1-\lambda t_2) + \frac{1}{2}\cos(\lambda t_1+\lambda t_2+2\theta)\right] \\
&= \frac{1}{2}\cos\lambda(t_1-t_2) \\
&\quad + \frac{1}{2}[\cos\lambda(t_1+t_2)E(\cos 2\theta) - \sin\lambda(t_1+t_2)E(\sin 2\theta)].
\end{aligned}
$$

当假定 θ 服从 $(-\pi, \pi)$ 上的均匀分布时，则此时 $\{\xi_t\}$ 为弱平稳过程．因为此时 $E(\cos 2\theta) = E(\sin 2\theta) = 0$，因此

$$
K(t_1, t_2) = \frac{1}{2}\cos\lambda(t_1-t_2).
$$

例 10.4.2 考虑一个具有随机振幅的正弦波，它由如下定义的随机过程描述．

$$
\xi_t = \xi\cos 2\pi t + \eta\sin 2\pi t, \ (t\in R_1).
$$

其中 ξ, η 为两个随机变数，且满足

$$
\begin{aligned}
E(\xi) &= E(\eta) = 0; \\
D(\xi) &= D(\eta) = 1; \\
E(\xi\eta) &= 0.
\end{aligned}
$$

则可计算得

$$
E(\xi_t) = \cos 2\pi t E(\xi) - \sin 2\pi t E(\eta) = 0,
$$

$$
\begin{aligned}
K(t_1, t_2) &= \cos 2\pi t_1\cos 2\pi t_2 E(\xi^2) + \sin 2\pi t_1\sin 2\pi t_2 E(\eta^2) \\
&\quad + \cos 2\pi t_1\sin 2\pi t_2 E(\xi\eta) + \sin 2\pi t_1\cos 2\pi t_2 E(\xi\eta) \\
&= \cos 2\pi(t_1-t_2).
\end{aligned}
$$

由此可得$\{\xi_t\}$是一个弱平稳过程.

二、平稳随机过程的相关函数

下面如无特别声明，平稳过程均指弱平稳过程而言. 由(10.4.2)式知,平稳过程的相关函数可用$B(\tau)$表示,即,

$$B(\tau)=E(\xi_{t+\tau}\cdot\xi_t). \qquad (10.4.3)$$

平稳过程的相关函数$B(\tau)$有如下性质:

性质 10.4.1

1) $B(0)\geqslant 0$;

2) $B(-\tau)=B(\tau)$;

3) $|B(\tau)|\leqslant B(0)$;

4) $B(\tau)$于R_1连续的充分必要条件为$B(\tau)$于$\tau=0$点连续;

5) $B(\tau)$是非负定的. 即，对任意$2n$个实数$a_1, a_2, \cdots, a_n$及$\tau_1, \tau_2, \cdots, \tau_n$,有

$$\sum_{i,\ j=1}^{n} B(\tau_i-\tau_j)a_i a_j\geqslant 0.$$

证 1),2)的证明可由(10.4.3)式立即得到.

现证3). 由于

$$0\leqslant E|\xi_{t+\tau}\pm\xi_t|^2=2[B(0)\pm B(\tau)],$$

因此得

$$B(0)\pm B(\tau)\geqslant 0.$$

即得

$$|B(\tau)|\leqslant B(0).$$

4) 的证明由下面的计算得到. 因为

$$\begin{aligned}
&|B(\tau+\Delta\tau)-B(\tau)|^2\\
=&|E(\xi_{\tau+\Delta\tau+\tau'}\xi_{\tau'})-E(\xi_{\tau+\Delta\tau+\tau'}\xi_{\Delta\tau+\tau'})|^2\\
=&|E[\xi_{\tau+\Delta\tau+\tau'}(\xi_{\tau'}-\xi_{\Delta\tau+\tau'})]|^2\\
\leqslant& E|\xi_{\tau+\Delta\tau+\tau'}|^2 E|\xi_{\tau'}-\xi_{\Delta\tau-\tau'}|^2\\
=&2B(0)[B(0)-B(\Delta\tau)]\to 0 \quad (当\ \Delta\tau\to 0).
\end{aligned}$$

因此，充分性得证．必要性是显然的．

5）的证明可从(10.1.9)直接得到．▌

读者可以证明，性质 10.4.1 中之 1)，2)，3)可直接从 5)得到．

到现在为止，我们考虑的随机过程均为实的随机过程，在理论研究和应用中，还须考虑复的随机过程．

$\{\xi_t\}$称为复的随机过程是指 ξ_t 可表为

$$\xi_t = X_t + jY_t,$$

其中 X_t, Y_t 均为实的随机过程．对于复随机过程的数学期望和相关函数定义如下：

$$E(\xi_t) = E(X_t) + jE(Y_t);$$

$$K(t_1, t_2) = E(\xi_{t_1}\bar{\xi}_{t_2}),\ (\bar{\xi}_{t_2}\text{表示 } \xi_{t_2}\text{的共轭复数}).$$

若复随机过程$\{\xi_t\}$满足下两式：

1) $E|\xi_t|^2 < \infty$;

2) $K(t+\tau, t) = B(\tau)$，(一般说，$B(\tau)$是实变复值函数)．

则称$\{\xi_t\}$为复的弱平稳过程．对于复的弱平稳过程，有类似于性质 10.4.1 的性质．现只列在下面，证明方法与性质 10.4.1 证明相仿，这里从略．

性质 10.4.2

1) $B(0) \geqslant 0$;

2) $B(-\tau) = \overline{B(\tau)}$;

3) $|B(\tau)| \leqslant B(0)$;

4) $B(\tau)$于 R_1 上连续的充分必要条件为 $B(\tau)$于 $\tau = 0$ 点连续；

5) $B(\tau)$为非负定的．即对任意 n 个复数 $z_1, z_2, \cdots, z_n$及 n 个实数 $\tau_1, \tau_2, \cdots, \tau_n$，有

$$\sum_{i,j=1}^{n} B(\tau_i - \tau_j) z_i \bar{z}_j \geqslant 0.$$

在通讯、控制、滤波理论和自然现象中，还必须考虑两个或两个以上的随机过程．例如在通讯中，信号在传输过程中受各种随机干扰的影响，接收端收到的是信号和噪声的叠加，信号和噪声一般来说都是随机过程．对两个随机过程，我们定义一个描述它们之间相关程度的数字特征，称为互相关函数．

定义 10.4.3 设$\{\xi_t\}$，$\{\eta_t\}$是两个随机过程（复的或实的）．称

$$K_{\xi,\eta}(t_1,t_2)=E(\xi_{t_1}\bar{\eta}_{t_2})\qquad(t_1,t_2\in R_1)$$

为$\{\xi_t\}$与$\{\eta_t\}$的互相关函数．

若对任意$t,\tau\in R_1$，

$$K_{\xi,\eta}(t+\tau,t)=B_{\xi,\eta}(\tau),$$

则称$\{\xi_t\}$与$\{\eta_t\}$平稳相关．

对于平稳相关的两个随机过程，其相关函数$B_{\xi,\eta}(\tau)$有如下两个性质：

1) $\overline{B_{\xi,\eta}(\tau)}=B_{\eta,\xi}(-\tau)$;

2) $|B_{\xi,\eta}(\tau)|^2\leqslant B_\xi(0)B_\eta(0)$.

性质 10.4.3 设随机过程$\{\xi_t\}$定义如下：

$$\xi_t=\eta_t+\zeta_t$$

其中$\{\eta_t\}$及$\{\zeta_t\}$为平稳相关的平稳过程（复的或实的），则$\{\xi_t\}$也是平稳过程，且其相关函数为

$$B_\xi(\tau)=B_\eta(\tau)+B_\zeta(\tau)+B_{\eta,\zeta}(\tau)+B_{\zeta,\eta}(\tau),\qquad(10.4.4)$$

若再设$\{\eta_t\}$与$\{\zeta_t\}$正交（若对任意t_1，$t_2\in R_1$，$E(\eta_{t_1}\bar{\zeta}_{t_2})=0$，则称$\{\eta_t\}$与$\{\zeta_t\}$正交），则(10.4.4)式变为

$$B_\xi(\tau)=B_\eta(\tau)+B_\zeta(\tau).\qquad(10.4.5)$$

性质 10.4.3 的结果可直接从定义出发立即得到．

注：以上考虑了$T=R_1$的情形．事实上，对$T=[0,\infty)$，$T=\{0,1,2,\cdots\}$，$T=\{0,\pm1,\pm2,\cdots\}$等等，上述定义、定理及性质仍成立，只需把"实数τ"分别改为"$\tau\geqslant0$"，"非负整数τ"，"整数τ"等即可．

§10.5 均方微积分与随机微分方程

这一节我们引进一些概念与结果，为今后研究随机过程之用.

建立一般函数微积分理论，首先要定义序列收敛的意义. 建立随机过程的微积分理论也是如此. 但随机序列有多种不同的收敛意义：依概率收敛、以概率为 1 收敛、依分布收敛、均方收敛等. 本节着重讲在某些理论与应用中使用较多的一种收敛——均方收敛.

一、随机序列均方收敛

定义 10.5.1 设$\{\xi_n;\ n\in N\}$为一实随机序列(其中N为全体正整数集，以下同)，ξ为一随机变数(也可以为一常数)，若$E(\xi_n)^2<\infty$，$E(\xi)^2<\infty$，且

$$\lim_{n\to\infty}E|\xi_n-\xi|^2=0, \tag{10.5.1}$$

则称$\{\xi_n\}$均方收敛于ξ，称ξ为$\{\xi_n\}$的均方极限，记为

$$\underset{n\to\infty}{\text{l.i.m}}\,\xi_n=\xi,$$

或简写为

$$\text{l.i.m}\,\xi_n=\xi.$$

由车贝谢夫不等式立即可以得到：若$\{\xi_n\}$均方收敛于ξ，则$\{\xi_n\}$依概率收敛于ξ. 事实上，由车贝谢夫不等式：

$$P\{|\xi_n-\xi|\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E|\xi_n-\xi|^2,$$

立即得到所要证的结果.

定理 10.5.1 $\{\xi_n\}$均方收敛于ξ的充分必要条件为

$$\lim_{n,m\to\infty}E|\xi_n-\xi_m|^2=0. \tag{10.5.2}$$

证 必要性是明显的，事实上，因为

$$|\xi_n-\xi_m|^2=|\xi_n-\xi+\xi-\xi_m|^2$$
$$\leqslant2|\xi_n-\xi|^2+2|\xi_m-\xi|^2,$$

故

$$0 \leqslant E|\xi_n - \xi_m|^2 \leqslant 2E|\xi_n - \xi|^2 + 2E|\xi_m - \xi|^2$$
$$\longrightarrow 0, \ (n, m \to \infty).$$

充分性的证明可参阅[23],[31].▌

定理 10.5.2 设$\{\xi_n\}$, $\{\eta_n\}$, $\{\zeta_n\}$为三个实随机序列, z 为随机变数, 皆有有限二阶矩. $\{c_n\}$是一常数序列, a, b 是两个常数. 令 $\xi = \mathrm{l.i.m}\,\xi_n$, $\eta = \mathrm{l.i.m}\,\eta_n$, $\zeta = \mathrm{l.i.m}\,\zeta_n$, $c = \lim\limits_{n\to\infty} c_n$. 则

(1) $\mathrm{l.i.m}\,c_n = \lim\limits_{n\to\infty} c_n = c$;

(2) $\mathrm{l.i.m}\,z = z$;

(3) $\mathrm{l.i.m}\,c_n z = cz$;

(4) $\mathrm{l.i.m}(a\xi_n + b\eta_n) = a\xi + b\eta$;

(5) $\lim\limits_{n\to\infty} E(\xi_n) = E(\xi)$, 即

$$E(\mathrm{l.i.m}\,\xi_n) = \lim_{n\to\infty} E(\xi_n);$$

(6) $\lim\limits_{n, m\to\infty} E(\xi_n \eta_m) = E(\xi\eta)$, 特别有

$$\lim_{n\to\infty} E(\xi_n^2) = E(\xi^2);$$

(7) 若 $E(\xi_n \eta_n) = E(\zeta_n)$, 则 $E(\xi\eta) = E(\zeta)$.

证 (1),(2),(3)由均方收敛的定义立即可得. (4)的证明则由下面的不等式得到:

$$E|a\xi_n + b\eta_n - (a\xi + b\eta)|^2 = E|a(\xi_n - \xi) + b(\eta_n - \eta)|^2$$
$$\leqslant 2a^2 E|\xi_n - \xi|^2 + 2b^2 E|\eta_n - \eta|^2.$$

(5) 的证明由许瓦兹不等式可得. 即由

$$0 \leqslant |E(\xi_n) - E(\xi)|^2 = |E(\xi_n - \xi)|^2$$
$$\leqslant E|\xi_n - \xi|^2 \to 0 \quad (n \to \infty)$$

可得(5).

现证(6)．因为，

$$|E(\xi_n\eta_m)-E(\xi\eta)|$$
$$=|E[(\eta-\eta_m)\xi]+E[(\xi-\xi_n)\eta]-E(\xi-\xi_n)(\eta-\eta_m)|$$
$$\leqslant|E(\eta-\eta_m)\xi|+|E(\xi-\xi_n)\eta|+|E(\xi-\xi_n)(\eta-\eta_m)|$$
$$\leqslant[E|\eta-\eta_m|^2E|\xi|^2]^{\frac{1}{2}}+[E|\xi-\xi_n|^2E|\eta|^2]^{\frac{1}{2}}$$
$$+[E|\xi-\xi_n|^2E|\eta-\eta_m|^2]^{\frac{1}{2}}\to 0,\ (n,m\to\infty).$$

(7)的证明如下：

由(5)，(6)及假设可得：

1° $E(\xi_n\eta_n)=E(\zeta_n)\to E(\zeta)$；

2° $E(\xi_n\eta_n)\to E(\xi\eta)$．

由1°及2°得$E(\xi\eta)=E(\zeta)$．∎

二、随机过程的均方连续

如无特别声明，以下都假定，随机过程是实的，其指标集 $T=R_1$，且有有限二阶矩．

定义 10.5.2 设$\{\xi_t\}$为一随机过程，若对 $t\in R_1$，

$$\underset{h\to 0}{\text{l.i.m}}\,\xi_{t+h}=\xi_t\ (\text{即}\lim_{h\to 0}E|\xi_{t+h}-\xi_t|^2=0),$$

则称$\{\xi_t\}$在 t 点均方连续．若对 R_1 中的每一点都均方连续，则称$\{\xi_t\}$在 R_1 上均方连续．

定理 10.5.3 $\{\xi_t\}$在 t 点均方连续的充分必要条件为其相关函数 $K(t_1,t_2)$在点(t,t)连续．

证 先证充分性．因为

$$E|\xi_{t+h}-\xi_t|^2=E(\xi_{t+h}\xi_{t+h})-E(\xi_{t+h}\xi_t)$$
$$-E(\xi_t\xi_{t+h})+E(\xi_t\xi_t)$$
$$=K(t+h,t+h)-K(t+h,t)-K(t,t+h)+K(t,t)$$
$$\to 0\ (h\to 0).$$

故充分性得证．

必要性的证明，可由定理 10.5.2 之(6)得到：

$$\lim_{h,h'\to 0} K(t+h, t+h') = \lim_{h,h'\to 0} E(\xi_{t+h}\xi_{t+h'})$$
$$= E(\xi_t\xi_t) = K(t,t). \blacksquare$$

推论 $\{\xi_t\}$的相关函数 $K(t_1, t_2)$在整个平面上每一点(t_1, t_2)连续的充要条件为它在对角线 $t_1 = t_2$ 上连续.

证 必要性是显然的，现证充分性. 由定理 10.5.3 知 $\{\xi_t\}$在每一点 $t\in R_1$ 均方连续，即若$t_1\in R_1$，$t_2\in R_1$，有

$$\underset{h\to 0}{\text{l.i.m}}\,\xi_{t_1+h} = \xi_{t_1}$$
$$\underset{h'\to 0}{\text{l.i.m}}\,\xi_{t_2+h'} = \xi_{t_2}$$

再由定理 10.5.2 之(6)得

$$\lim_{h,h'\to 0} K(t_1+h, t_2+h') = \lim_{h,h'\to 0} E(\xi_{t_1+h}\xi_{t_2+h'})$$
$$= E(\xi_{t_1}\xi_{t_2})$$
$$= K(t_1, t_2). \blacksquare$$

定理 10.5.3 说明随机过程的均方连续性与它的相关函数的连续性是等价的，而定理 10.5.3 的推论则指出了相关函数关于连续性的一个有趣事实：只需要相关函数在对角线 $t_1 = t_2$ 的每一点上连续，则这个相关函数必在整个平面 $R_1\times R_1$ 上连续.

三、随机过程的均方积分

设$\{\xi_t\}$为一随机过程， $a, b\in R_1$，$a<b$. 将区间$[a, b]$作分割如下：

$$a = t_0 < t_1 < \cdots < t_n = b.$$

又

$$\Delta = \max_{0\leqslant i\leqslant n-1}\{(t_{i+1} - t_i)\},$$
$$t_i \leqslant t_i' < t_{i+1}\ (i = 0, 1, \cdots, n-1).$$

定义 10.5.3 若按照上述记号，和式

$$\sum_{i=0}^{n-1} \xi_{t_i'}(t_{i+1}-t_i)$$

在均方意义下存在有限极限，则说$\{\xi_t\}$在区间$[a, b]$上均方黎曼可积分(或简称均方可积). 其极限值称为$\{\xi_t\}$在区间$[a, b]$上的均方积分，并以$\int_a^b \xi_t dt$表示，即

$$\lim_{\Delta\to 0} E\left|\sum_{i=0}^{n-1} \xi_{t_i'}(t_{i+1}-t_i)-\int_a^b \xi_t dt\right|^2=0,$$

或

$$\underset{\Delta\to 0}{\mathrm{l.i.m}}\sum_{i=0}^{n-1} \xi_{t_i'}(t_{i+1}-t_i)=\int_a^b \xi_t dt.$$

由定理 10.5.2 之(5)立即得到，若$\{\xi_t\}$在$[a, b]$上均方可积，则有

$$E\left[\int_a^b \xi_t dt\right]=\int_a^b E(\xi_t) dt,$$

即求数学期望和求均方积分可调换次序.

定理 10.5.4 随机过程$\{\xi_t\}$在$[a, b]$上均方可积的充分条件为$\{\xi_t\}$的相关函数$K(t_1, t_2)$在矩形$[a,b]\times[a,b]$上依普通意义下黎曼可积.

证 令对区间$[a,b]$的另一分划为

$$a\leqslant \tilde{t}_0<\tilde{t}_1<\cdots<\tilde{t}_m=b.$$

又令

$$\tilde{\Delta}=\max_{0\leqslant k\leqslant m-1}\{(\tilde{t}_{k+1}-\tilde{t}_k)\},$$

$$\tilde{t}_k\leqslant \tilde{t}_k'<\tilde{t}_{k+1}\quad (k=0,1,\cdots m-1).$$

只需证明，当$\Delta, \tilde{\Delta}\to 0$时，

$$E\left|\sum_{i=0}^{n-1}\xi_{t'_i}(t_{i+1}-t_i)-\sum_{k=0}^{m-1}\xi_{\tilde{t}'_k}(\tilde{t}_{k+1}-\tilde{t}_k)\right|^2\to 0.$$

现因为

$$\begin{aligned}
&E\left|\sum_{i=0}^{n-1}\xi_{t'_i}(t_{i+1}-t_i)-\sum_{k=0}^{m-1}\xi_{\tilde{t}'_k}(\tilde{t}_{k+1}-\tilde{t}_k)\right|^2\\
=&E\left[\sum_{i=0}^{n-1}\xi_{t'_i}(t_{i+1}-t_i)\cdot\sum_{j=0}^{n-1}\xi_{t'_j}(t_{j+1}-t_j)\right]\\
&+E\left[\sum_{j=0}^{m-1}\xi_{\tilde{t}'_j}(\tilde{t}_{j+1}-\tilde{t}_j)\cdot\sum_{k=0}^{m-1}\xi_{\tilde{t}'_k}(\tilde{t}_{k+1}-\tilde{t}_k)\right]\\
&-2E\left[\sum_{i=0}^{n-1}\xi_{t'_i}(t_{i+1}-t_i)\cdot\sum_{k=0}^{m-1}\xi_{\tilde{t}'_k}(\tilde{t}_{k+1}-\tilde{t}_k)\right]\\
=&\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}E(\xi_{t'_i}\xi_{t'_j})(t_{i+1}-t_i)(t_{j+1}-t_j)\\
&+\sum_{j=0}^{m-1}\sum_{k=0}^{m-1}E(\xi_{\tilde{t}'_j}\xi_{\tilde{t}'_k})(\tilde{t}_{j+1}-\tilde{t}_j)(\tilde{t}_{k+1}-\tilde{t}_k)\\
&-2\sum_{i=0}^{n-1}\sum_{k=0}^{m-1}E(\xi_{t'_i}\xi_{\tilde{t}'_k})(t_{i+1}-t_i)(\tilde{t}_{k+1}-\tilde{t}_k),
\end{aligned}$$

由此式可见，当 $K(t_1,t_2)$ 于 $[a,b]\times[a,b]$ 黎曼可积时，上式右边当 $\Delta,\tilde{\Delta}\to 0$ 时趋于0，再由定理 10.5.1 及定义 10.5.3 可得 $\{\xi_t\}$ 均方可积. ▌

推论 若 $\{\xi_t\}$ 于 $[a,b]$ 均方连续，则必均方可积.

证 由假设及定理 10.5.3 可得 $\{\xi_t\}$ 的相关函数 $K(t_1,t_2)$ 在 $[a,b]\times[a,b]$ 上连续，故 $K(t_1,t_2)$ 于 $[a,b]\times[a,b]$ 黎曼可积，再由定理 10.5.4 即得证. ▌

对于均方连续于区间 $[a,b]$ 的随机过程，其均方积分有如下性质（由于证明方法与普通数学分析中关于积分性质的证明方法类

似，读者可自行补足其证明）：

性质10.5.1 设随机过程$\{\xi_t\}$与$\{\eta_t\}$皆于$[a,b]$均方连续，则：

（1）成立不等式：

$$E\left|\int_a^b \xi_t dt\right|^2 \leqslant M(b-a)^2$$

其中

$$M=\max_{a\leqslant t\leqslant b} E|\xi_t|^2.$$

（2）对$a<b<c$，有

$$\int_a^c \xi_t dt=\int_a^c \xi_t dt+\int_c^b \xi_t dt;$$

（3）对常数，α,β，有

$$\int_a^b (\alpha\xi_t+\beta\eta_t)dt=\alpha\int_a^b \xi_t dt+\beta\int_a^b \eta_t dt.$$

下面我们来考虑随机过程$\{\xi_t\}$的不定积分.

若$\{\xi_t\}$于$[a,b]$上均方连续，令

$$\eta_t=\int_a^t \xi_s ds.$$

称η_t为ξ_t的不定积分（在均方意义下）. 由上述定理10.5.4的推论可知η_t对任意$t\in[a,b]$总是存在的. 我们可以证明下面的定理成立.

定理10.5.5 若随机过程$\{\xi_t\}$于$[a,b]$均方连续，则它的不定积分η_t所构成的随机过程$\{\eta_t\}$在$[a,b]$上均方连续.

证 由性质10.5.1得，

$$E|\eta_{t+h}-\eta_t|^2=E\left|\int_a^{t+h}\xi_s ds-\int_a^t \xi_s ds\right|^2$$

$$=E\left|\int_t^{t+h}\xi_s ds\right|^2$$

$$\leqslant h^2\max_{a\leqslant t\leqslant b}E|\xi_t|^2\to 0(\text{当 } h\to 0). \blacksquare$$

四、随机过程的均方导数

定义 10.5.4 设$\{\xi_t\}$为一随机过程，若存在随机过程$\{\xi'_t\}$，使得对 $t\in R_1$，

$$\underset{h\to 0}{\text{l.i.m}}\left(\frac{\xi_{t+h}-\xi_t}{h}\right)=\xi'_t,$$

则称$\{\xi_t\}$于 t 点均方可微，且称ξ'_t为$\{\xi_t\}$在 t 点的均方导数．若对每一 $t\in R_1$，$\{\xi_t\}$都均方可微，则称$\{\xi_t\}$在 R_1 上均方可微．

定理 10.5.6 随机过程$\{\xi_t\}$在 R_1 上均方可微的充分必要条件为它的相关函数 $K(t_1,t_2)$的广义二阶导数$\frac{\partial^2 K(t_1,t_2)}{\partial t_1\partial t_2}$①对于对角线上的每一点$(t,t)\in R_1\times R_1$上存在．

证 先证充分性．因为

$$\begin{aligned}&E\left|\frac{\xi_{t+h}-\xi_t}{h}-\frac{\xi_{t+h'}-\xi_t}{h'}\right|^2\\&=E\left|\frac{\xi_{t+h}-\xi_t}{h}\right|^2+E\left|\frac{\xi_{t+h'}-\xi_t}{h'}\right|^2\\&\quad-2E\left(\frac{\xi_{t+h}-\xi_t}{h}\cdot\frac{\xi_{t+h'}-\xi_t}{h'}\right).\end{aligned}$$

由$\frac{\partial^2 K(t_1,t_2)}{\partial t_1\partial t_2}$于$(t,t)$存在这一假设得

$$\begin{aligned}&E\left(\frac{\xi_{t+h}-\xi_t}{h}\cdot\frac{\xi_{t+h'}-\xi_t}{h'}\right)\\&=\frac{1}{hh'}[K(t+h,t+h')-K(t,t+h')\\&\quad-K(t+h,t)+K(t,t)]\end{aligned}$$

① 若二元函数 $f(x_1,x_2)$有如下极限存在：

$$\lim_{\substack{h\to 0\\h'\to 0}}\frac{f(x_1+h,x_2+h')-f(x_1+h,x_2)-f(x_1,x_2+h')+f(x_1,x_2)}{hh'},$$

则说 $f(x_1,x_2)$在点(x_1,x_2)的广义二阶导数存在，且称极限为 $f(x_1,x_2)$在点(x_1,x_2)的广义二阶导数．

$$\to \frac{\partial^2 K(t,t)}{\partial t \partial t} \quad (h, h' \to 0)$$

同理可得

$$E\left|\frac{\xi_{t+h}-\xi_t}{h}\right|^2 = E\left(\frac{\xi_{t+h}-\xi_t}{h}\right)\left(\frac{\xi_{t+h}-\xi_t}{h}\right)$$

$$\to \frac{\partial^2 K(t,t)}{\partial t \partial t} \quad (h \to 0)$$

及

$$E\left|\frac{\xi_{t+h'}-\xi_t}{h'}\right|^2 = E\left(\frac{\xi_{t+h'}-\xi_t}{h'}\right)\left(\frac{\xi_{t+h'}-\xi_t}{h'}\right)$$

$$\to \frac{\partial^2 K(t,t)}{\partial t \partial t} \quad (h' \to 0).$$

因此可得

$$E\left|\frac{\xi_{t+h}-\xi_t}{h} - \frac{\xi_{t+h'}-\xi_t}{h'}\right|^2 \to 0 \, (h,\ h' \to 0).$$

根据定理 10.5.1 即可得$\{\xi_t\}$于 $t \in R_1$ 可微.

必要性证明如下:

设$\{\xi_t\}$于 $t \in R_1$ 的导数为 ξ'_t，则由定理 10.5.2 中之(6)有

$$\lim_{h, h' \to 0} E\left(\frac{\xi_{t+h}-\xi_t}{h} \frac{\xi_{t+h'}-\xi_t}{h'}\right) = E(\xi'_t \xi'_t).$$

由充分性的证明过程知，上式左端等于

$$\lim_{h, h' \to 0} \frac{1}{hh'}[K(t+h, t+h') - K(t, t+h') - K(t+h, t) + K(t,t)].$$

因此可得

$$\lim_{h, h' \to 0} \frac{1}{hh'}[K(t+h, t+h') - K(t, t+h')$$

$$- K(t+h, t) + K(t,t)] = E(\xi'_t \xi'_t).$$

即 $K(t_1, t_2)$ 于$(t,\ t) \in R_1 \times R_1$ 的广义二阶导数$\frac{\partial^2 K(t,t)}{\partial t \partial t}$存在.

推论 若$\frac{\partial^2 K(t_1,t_2)}{\partial t_1 \partial t_2}$在对角线点$(t,t)\in R_1\times R_1$存在，则对任意$(t_1,t_2)\in R_1\times R_1$存在，且

$$\frac{\partial^2 K(t_1,t_2)}{\partial t_1 \partial t_2}=E(\xi'_{t_1}\xi'_{t_2}).$$

证 由题设及定理 10.5.6 得知，$\{\xi_t\}$于$t\in R_1$上可微，即对$t_1,t_2\in R_1$，有

$$\underset{h\to 0}{\text{l.i.m}}\frac{\xi_{t_1+h}-\xi_{t_1}}{h}=\xi'_{t_1},$$

$$\underset{h'\to 0}{\text{l.i.m}}\frac{\xi_{t_2+h'}-\xi_{t_2}}{h'}=\xi'_{t_2}.$$

再由定理 10.5.2 中之(6)可得

$$E(\xi'_{t_1}\xi'_{t_2})=\lim_{h,h'\to 0}E\left(\frac{\xi_{t_1+h}-\xi_{t_1}}{h}\frac{\xi_{t_2+h'}-\xi_{t_2}}{h'}\right)$$

$$=\frac{\partial^2 K(t_1,t_2)}{\partial t_1 \partial t_2}.\ \blacksquare$$

均方导数有很多类似普通函数的导数的性质，例如，

性质 10.5.2

(1) 任一随机变数ξ(可以是常数)的均方导数为 0；

(2) $(\xi_t+\xi)'=\xi'_t$；由此可得，若两个随机过程的均方导数相等，则它们相差一个随机变数(可以是常数).

(3) 若$\{\xi_t\}$与$\{\eta_t\}$于$t\in R_1$均方可微，α,β为常数，那么$\{\alpha\xi_t+\beta\eta_t\}$于$t\in R_1$均方可微，且

$$(\alpha\xi_t+\beta\eta_t)'=\alpha\xi'_t+\beta\eta'_t.$$

定理 10.5.7 若$\{\xi_t\}$于R_1均方可微，则对任意t_1，$t_2\in R_1$有

(1) $\frac{d}{dt_1}E(\xi_{t_1})=E(\xi'_{t_1})$；

(2) $\frac{\partial K(t_1,t_2)}{\partial t_1}=\frac{\partial}{\partial t_1}E(\xi_{t_1}\xi_{t_2})=E(\xi'_{t_1}\xi_{t_2})$；

(3) $\frac{\partial^2 K(t_1, t_2)}{\partial t_1 \partial t_2} = \frac{\partial}{\partial t_1} \frac{\partial}{\partial t_2} E(\xi_{t_1} \xi_{t_2}) = E(\xi'_{t_1} \xi'_{t_2})$.

证

(1) 由定理 10.5.2 之(5)得

$$E(\xi'_{t_1}) = \lim_{h \to 0} E\left(\frac{\xi_{t_1+h} - \xi_{t_1}}{h}\right) = \lim_{h \to 0} \frac{1}{h}[E(\xi_{t_1+h}) - E(\xi_{t_1})]$$

$$= [E(\xi_{t_1})]',$$

即

$$\frac{d}{dt_1} E(\xi_{t_1}) = E(\xi'_{t_1}).$$

(2) 因为

$$\frac{\xi_{t_1+h}\xi_{t_2} - \xi_{t_1}\xi_{t_2}}{h} = \frac{\xi_{t_1+h} - \xi_{t_1}}{h} \xi_{t_2},$$

由定理 10.5.2 之(6)得

$$\lim_{h \to 0} E\left(\frac{\xi_{t_1+h} - \xi_{t_1}}{h} \xi_{t_2}\right) = E(\xi'_{t_1} \xi_{t_2}),$$

而上式左端等于

$$\lim_{h \to 0} \frac{1}{h}[K(t_1 + h, t_2) - K(t_1, t_2)].$$

因而得

$$\frac{\partial K(t_1, t_2)}{\partial t_1} = E(\xi'_{t_1} \xi_{t_2}).$$

(3) 的证明同上，请读者自行补足. ▌

定理 10.5.7 说明，当$\{\xi_t\}$在 R_1 上均方可微时，则求数学期望与求微商可交换次序.

定理 10.5.8 设随机过程$\{\xi_t\}$于区间$[a, b]$上均方连续，则它的不定积分

$$\eta_t = \int_a^t \xi_s ds, \quad t \in [a, b]$$

在$[a,b]$上均方可微，且

$$\eta'_t = \xi_t$$

证 由假设及性质 10.5.1 可得

$$E\left|\frac{\eta_{t+h}-\eta_t}{h}-\xi_t\right|^2 = E\left|\frac{1}{h}\int_t^{t+h}\xi_s ds-\xi_t\right|^2$$

$$=\frac{1}{h^2}E\left|\int_t^{t+h}(\xi_s-\xi_t)ds\right|^2$$

$$\leqslant \max_{t\leqslant s\leqslant t+h}E\ |\xi_s-\xi_t|^2\to 0 \quad (h\to 0).$$

由此得到$\{\eta_t\}$于$[a,b]$均方可微，且对任意 $t\in[a,b]$

$$\eta'_t = \xi_t. \ ▮$$

推论 设$\{\xi_t\}$于$[a,b]$均方可微，且$\{\xi'_t\}$均方连续，则对任意$t\in[a,b]$，有

$$\xi_t-\xi_a=\int_a^t \xi'_s ds,$$

特别当 $t=b$ 时，有

$$\xi_b-\xi_a=\int_a^b \xi'_s ds.$$

证 事实上，由定理 10.5.8 及性质 10.5.1 和 性 质 10.5.2 即得. ▮

五、随机微分方程

利用分析数学来描述某些自然现象时，通常要建立一个微分方程，并给定初始条件或边界条件. 例如描述一个线性系统输入与输出关系可有如下微分方程及初始条件：

$$a_n y^{(n)}+\cdots+a_1 y'+a_0 y=x(t),$$
$$y(0)=b.$$

但考虑到随机因素的影响，例如初始条件的轻微改变、由测量误差带来的常系数改变或$x(t)$本身是一随机过程（如$x(t)$表示作布朗运动的质点受到随机碰撞）. 这都使得方程的解 $y(t)$ 不可能是

一决定性函数而为一随机过程．这些事实促使建立了近代概率论的一个重要分支——随机微分方程．

考虑一个微分方程

$$a_n\xi_t^{(n)}+a_{n-1}\xi_t^{(n-1)}+\cdots+a_1\xi_t'+a_0\xi_t=\eta_t, \qquad (10.5.3)$$

若 $a_k(k=0,1,\cdots n)$ 是随机变数，$\{\xi_t\}$，$\{\eta_t\}$ 是随机过程，则称上述微分方程为随机微分方程．自然，上述之 a_k，$\{\xi_t\}$ 或 $\{\eta_t\}$ 可以有一个或一个以上是普通的常数或函数，假若全部都是普通的常数或函数，则(10.5.3)式就是熟知的微分方程．

随机微分方程可以用来描述一个系统的输入与输出关系，若把 $\{\xi_t\}$ 看作是一个输入信号（它是随机的），则 $\{\eta_t\}$ 可视为通过该系统的输出．因此，在控制论、滤波及通讯理论方面都需要用(10.5.3)式所表达的随机微分方程来作为重要的数学工具．现在随机微分方程不论在理论方面还是在应用方面都得到了迅速的发展，读者如需较系统的学习，可参看比较基本的教材，如[23]，[33]．我们在下面只介绍一个简单例子，建立一些概念．

在(10.5.3)式中，我们考虑当 $n=1$，a_0 及 a_1 皆为常数的情形，即考虑一阶常系数随机微分方程：

$$a_1\xi_t'+a_0\xi_t=\eta_t \qquad t\in[0,\infty), \qquad (10.5.4)$$

给定初始条件

$$\xi_0=0 \ (\text{以概率为 } 1).$$

下面我们讨论 $\{\xi_t\}$ 与 $\{\eta_t\}$ 之间的一些特征的关系（这里主要介绍数学期望与相关函数），假定它们的二阶矩都存在．

1) $\{\xi_t\}$ 与 $\{\eta_t\}$ 的数学期望的关系．

对(10.5.4)两端求数学期望，并根据定理 10.5.7 之(1)得

$$\begin{aligned}E(\eta_t)&=a_1E(\xi_t')+a_0E(\xi_t)\\&=a_1E'(\xi_t)+a_0E(\xi_t).\end{aligned} \qquad (10.5.5)$$

此时，初始条件为

$$E(\xi_0)=0$$

(10.5.5)式是普通的一阶常微分方程，由熟知的解法，可得$E(\xi_t)$与$E(\eta_t)$的关系．换句话说，我们若能知道输入$\{\xi_t\}$的数学期望，则可通过解(10.5.5)式得到输出$\{\eta_t\}$的数学期望．反之，已知$E(\eta_t)$，可求$E(\xi_t)$.

2) $\{\xi_t\}$与$\{\eta_t\}$的相关函数的关系.

设$t_1, t_2\in[0,\infty)$，由(10.5.4)式:

$$a_1\xi'_{t_2}+a_0\xi_{t_2}=\eta_{t_2}$$

两边乘以η_{t_1}得

$$a_1\eta_{t_1}\xi'_{t_2}+a_0\eta_{t_1}\xi_{t_2}=\eta_{t_1}\eta_{t_2},$$

上式两边求数学期望得

$$a_1E(\eta_{t_1}\xi'_{t_2})+a_0E(\eta_{t_1}\xi_{t_2})=E(\eta_{t_1}\eta_{t_2}).$$

由定理10.5.7之(2)，上式为

$$a_1\frac{\partial K_{\eta,\xi}(t_1,t_2)}{\partial t_2}+a_0K_{\eta,\xi}(t_1,t_2)=K_\eta(t_1,t_2). \quad (10.5.6)$$

此时的初始条件为

$$K_{\eta,\xi}(t_1,0)=0.$$

用同样方法，若对(10.5.4)式令$t=t_1$，且两端乘以ξ_{t_2}，也可得到类似的常微分方程:

$$a_1\frac{\partial K_\xi(t_1,t_2)}{\partial t_1}+a_0K_\xi(t_1,t_2)=K_{\eta,\xi}(t_1,t_2) \qquad (10.5.7)$$

初始条件为

$$K_\xi(0,t_2)=0.$$

我们联立方程(10.5.6)及(10.5.7)，则可解得$\{\xi_t\}$与$\{\eta_t\}$的相关函数之间的关系.

例 10.5.1 考虑随机微分方程

$$\begin{cases}\xi'_t+\alpha\xi_t=\eta_t & t\in[0,\infty),\\ \xi_0=0,\end{cases} \qquad (10.5.8)$$

其中 $\alpha>0$，$\{\eta_t\}$为弱平稳过程，且

$$E(\eta_t)=\lambda, B_\eta(\tau)=e^{-2\lambda|\tau|}, \qquad (10.5.9)$$

其中 $\lambda>0$.

由(10.5.5)，(10.5.8)两式及上式可得

$$\begin{cases}E'(\xi_t)+\alpha E(\xi_t)=\lambda,\\ E(\xi_0)=0.\end{cases}$$

解上述微分方程得

$$E(\xi_t)=\frac{\lambda}{\alpha}(1-e^{-\alpha t}).$$

$E(\xi_t)$ 与 $E(\eta_t)$ 的关系如图 10.5.1 所示

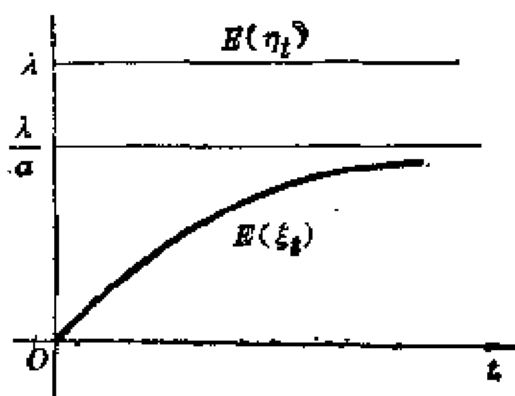


图 10.5.1

由(10.5.6)及(10.5.9)两式得

$$\frac{\partial K_{\eta\xi}(t_1,t_2)}{\partial t_2}+\alpha K_{\eta\xi}(t_1,t_2)=e^{-2\lambda|t_1-t_2|}$$

此时的初始条件由(10.5.8)式可知为

$$K_{\eta\xi}(t_1,0)=0.$$

因而可解得当 $t_1>t_2$ 时

$$K_{\eta\xi}(t_1,t_2)=\frac{1}{2\lambda+\alpha}e^{-2\lambda t_1}[e^{2\lambda t_2}-e^{-\alpha t_2}],$$

将上式代入(10.5.7)式得

$$\frac{\partial K_\xi(t_1,t_2)}{\partial t_1}+\alpha K_\xi(t_1,t_2)=\frac{1}{2\lambda+\alpha}e^{-2\lambda t_1}[e^{2\lambda t_2}-e^{-\alpha t_2}].$$

此时初始条件为 $K_\xi(0,t_2)=0$，由此解得

$$K_\xi(t_1,t_2)=\frac{1}{(\alpha+2\lambda)(\alpha-2\lambda)}(e^{2\lambda t_2}-e^{-\alpha t_2})(e^{-2\lambda t_1}-e^{-\alpha t_1}).$$

对于 $t\leqslant t_2$时，可类似求解.

3) 随机微分方程的解.

下面只给出一个例子.

例 10.5.2 设随机微分方程为

$$\xi_t' = \eta_t \qquad t \in [a, b], \qquad (10.5.10)$$

初始条件为

$$\xi_a = \xi_0,$$

其中$\{\eta_t\}$于$[a,b]$均方可积，ξ_0是数学期望有限的随机变数.

由定理 10.5.8 的推论得，当$t\in[a,b]$，有

$$\xi_t - \xi_a = \int_a^t \eta_s ds,$$

由初始条件得

$$\xi_t = \xi_0 + \int_a^t \eta_s ds.$$

$\{\xi_t\}$的数学期望为

$$E(\xi_t) = E(\xi_0) + \int_a^t E(\eta_s) ds.$$

如若设ξ_0服从正态$N(0,1)$，且$\{\eta_t\}$为例 10.5.1 所述的弱平稳过程，则有

$$E(\xi_t) = \int_a^t \lambda ds = \lambda(t-a).$$

§ 10.6 弱平稳随机过程的功率谱密度

在无线电技术和其它应用问题里，特别是在通讯技术中，经常要考虑所谓功率、能量、功率谱密度. 我们知道，表示一个信号的函数（例如电压的大小，用$f(t)$表示）可以定义它的功率、平均功率等. 由于在我们所讨论的问题中，用于描述信号的函数是一个随机过程，我们需要类似地引入上述概念，下面我们只考虑关于功率谱概念.

定义 10.6.1 设$\{\xi_t\}$为一弱平稳过程，$B(\tau)$为其相关函数. 当$B(\tau)$的富氏变换

$$\int_{-\infty}^{\infty} e^{-j\omega\tau} B(\tau) d\tau^{①}$$

存在时，我们称它为$\{\xi_t\}$的谱密度，并记为 $S(\omega)$. 即

$$S(\omega) = \int_{-\infty}^{\infty} e^{-j\omega\tau} B(\tau) d\tau. \qquad (10.6.1)$$

$S(\omega)$的意义，我们将在稍后加以解释.

例 10.6.1 设$\{\xi_t\}$为例 10.1.7 所定义的随机过程，已知其相关函数为

$$B(\tau) = e^{-2\lambda|\tau|}.$$

根据(10.6.1)式可算得$\{\xi_t\}$的谱密度为

$$S(\omega) = \int_{-\infty}^{\infty} e^{-2\lambda|\tau|} e^{-j\omega\tau} d\tau = \frac{4\lambda}{4\lambda^2 + \omega^2}.$$

$\{\xi_t\}$的相关函数与谱密度如图 10.6.1 之(a)及(b).

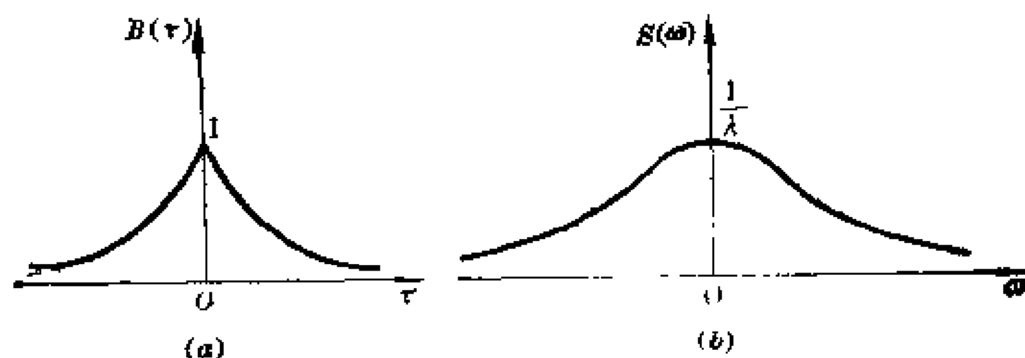


图 10.6.1

性质 10.6.1 谱密度 $S(\omega)$有如下性质:

1) $S(\omega)$是 ω 的实函数;

2) 当$\{\xi_t\}$为实随机过程时，$S(\omega)$关于 O 点对称，且

$$S(\omega) = 2\int_{0}^{\infty} B(\tau) \cos\omega\tau d\tau;$$

---

① 这里的参数ω不是概率空间$(\Omega, \mathscr{F}, P)$中的点，我们仅沿用许多书上的记号而已，学习时应加以注意.

3) 若弱平稳过程$\{\eta_t\}$与$\{\zeta_t\}$正交，则由 $\xi_t=\eta_t+\zeta_t$ 构成的随机过程$\{\xi_t\}$的谱密度为

$$S_\xi(\omega)=S_\eta(\omega)+S_\zeta(\omega),$$

其中 $S_\eta(\omega)$，$S_\zeta(\omega)$分别为$\{\eta_t\}$及$\{\zeta_t\}$的谱密度，皆假设存在。

证 1) 因为$\overline{B(\tau)}=B(-\tau)$，故

$$\overline{S(\omega)}=\int_{-\infty}^{\infty}\overline{B(\tau)}\,e^{j\omega\tau}d\tau=\int_{-\infty}^{\infty}B(-\tau)e^{j\omega\tau}d\tau$$

$$=\int_{-\infty}^{\infty}B(\tau)e^{-j\omega\tau}d\tau=S(\omega)$$

2) 用与1) 中同样的证明方法，可得

$$S(-\omega)=S(\omega).$$

并且由 1)及 $B(\tau)$对称于O点可得

$$S(\omega)=\int_{-\infty}^{\infty}B(\tau)e^{-j\omega\tau}d\tau=\int_{-\infty}^{\infty}B(\tau)\cos\omega\tau d\tau$$

$$=2\int_{0}^{\infty}B(\tau)\cos\omega\tau d\tau.$$

3) 由假设及(10. 4. 5)式，即可得所要结果. ∎

若$S(\omega)$满足一定的条件$\left(\text{如}\int_{-\infty}^{\infty}|S(\omega)|d\omega<\infty\right)$时，则有

$$B(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)e^{j\omega\tau}d\omega, \qquad (10.6.2)$$

上式当 $\tau=0$ 时，有

$$E|\xi_t|^2=B(0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)d\omega,$$

我们称 $E|\xi_t|^2$ 为$\{\xi_t\}$的平均功率，它等于$\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)d\omega$.

类似地，可以对两个平稳相关的弱平稳过程 $\{\xi_t\}$与$\{\eta_t\}$定义互谱密度如下：

$$S_{\xi\eta}(\omega)=\int_{-\infty}^{\infty}B_{\xi\eta}(\tau)e^{-j\omega\tau}d\tau$$

(当等式右边积分存在时).

若$\int_{-\infty}^{\infty}|S_{\xi,\eta}(\omega)|d\omega<\infty$时，有

$$B_{\xi,\eta}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S_{\xi,\eta}(\omega)e^{j\omega\tau}d\omega.$$

下面我们导出一些结果，说明$S(\omega)$的物理意义

定义 10.6.2 设$\{\xi_t\}$为一弱平稳过程，令

$$\tilde{S}_c(\omega)=\frac{1}{2c}\left|\int_{-c}^{c}\xi_t e^{-j\omega t}dt\right|^2 \text{①}\quad (c>0),$$

若极限

$$\lim_{c\to\infty}\frac{1}{2c}E\left|\int_{-c}^{c}\xi_t e^{-j\omega t}dt\right|^2$$

存在，则称它为$\{\xi_t\}$的功率谱密度，并以$\tilde{S}(\omega)$表示，即

$$\tilde{S}(\omega)=\lim_{c\to\infty}\frac{1}{2c}E\left|\int_{-c}^{c}\xi_t e^{-j\omega t}dt\right|^2. \qquad (10.6.3)$$

定理 10.6.1 若弱平稳过程$\{\xi_t\}$均方连续且它的相关函数绝对可积，即若

$$\int_{-\infty}^{\infty}|B(\tau)|d\tau<\infty,$$

则有

$$\tilde{S}(\omega)=S(\omega),$$

并且由(10.6.3)式可知$S(\omega)\geqslant 0$.

证 由定理 10.5.4 的推论可得

$$E[\tilde{S}_c(\omega)]=E\left[\frac{1}{2c}\left|\int_{-c}^{c}\xi_t e^{-j\omega t}dt\right|^2\right]$$

$$=\frac{1}{2c}\int_{-c}^{c}\int_{-c}^{c}E(\xi_{t_1}\bar{\xi}_{t_2})e^{-j\omega(t_1-t_2)}dt_1dt_2$$

---

① 随机积分$\int_{-c}^{c}\xi_t e^{-j\omega t}dt$的定义类似于定义 10.5.3，并且有类似的定理和性质.

$$=\frac{1}{2c}\int_{-c}^{c}\int_{-c}^{c}B(t_1-t_2)e^{-j\omega(t_1-t_2)}dt_1dt_2$$

$$=\frac{1}{2c}\int_{-2c}^{2c}(2c-|\tau|)B(\tau)e^{-j\omega\tau}d\tau$$

$$=\int_{-2c}^{2c}B(\tau)e^{-j\omega\tau}d\tau-\frac{1}{2c}\int_{-2c}^{2c}|\tau|B(\tau)e^{-j\omega\tau}d\tau.$$

其中,二重积分化为单积分时,用了变量替换: $t_1-t_2=\tau, t_2=t_2$.

若能证明上式右端第二项在假设条件下,当 $c\to\infty$ 时趋于 0,则定理 10.6.1 得证. 下面加以证明:

由假设, $\int_{-\infty}^{\infty}|B(\tau)|d\tau<\infty$,故任给 $\varepsilon>0$,存在$N_0>0$,使得当 $c>N_0$时

$$\int_{2c}^{\infty}|B(\tau)|d\tau\leqslant\int_{2N}^{\infty}|B(\tau)|d\tau<\frac{\varepsilon}{3};$$

$$\int_{-\infty}^{-2c}|B(\tau)|d\tau\leqslant\int_{-\infty}^{-2N_0}|B(\tau)|d\tau<\frac{\varepsilon}{3}.$$

将区间$[-2c,2c]$分成三个部分:

$$\Delta_1=[-2c,-2N_0);$$
$$\Delta_2=[-2N_0,2N_0);$$
$$\Delta_3=[2N_0,2c].$$

对于区间 Δ_1,有

$$\left|\frac{1}{2c}\int_{-2c}^{-2N_0}|\tau|B(\tau)e^{-j\omega\tau}d\tau\right|\leqslant\frac{2c}{2c}\int_{-2c}^{-2N_0}|B(\tau)|d\tau$$
$$\leqslant\int_{-\infty}^{-2N_0}|B(\tau)|d\tau$$
$$<\frac{\varepsilon}{3}. \tag{10.6.4}$$

同理,对区间 Δ_3,有

$$\left|\frac{1}{2c}\int_{2N_0}^{2c}|\tau|B(\tau)e^{-j\omega\tau}d\tau\right|<\frac{\varepsilon}{3}. \tag{10.6.5}$$

对于区间 Δ_2，有

$$\left|\frac{1}{2c}\int_{-2N_0}^{2N_0}|\tau|B(\tau)e^{-j\omega\tau}d\tau\right|\leqslant\frac{2N_0}{2c}\int_{-2N_0}^{2N_0}|B(\tau)|d\tau.$$

对于固定的 N_0，由假设，存在 N_1，当 $c>N_1$ 时，上式小于$\frac{\varepsilon}{3}$，由此及(10.6.4)，(10.6.5)两式，即得所要的结果. ▌

§10.7 遍历性定理

在实际工作中，确定随机过程的特征是很重要的. 例如飞机在高空飞行受湍流影响产生机翼振动，要确定其振动范围，以决定制造飞机时选取什么材料，这就要考虑机翼在时间变化过程中振动振幅大小(它是一随机过程)的方差. 又如电路中电子的不规则运动——热骚动引起的电位的脉动(通常称为热噪声，它也是随机过程)，需要考虑其脉动范围、噪声功率等，也要考虑其方差、相关函数等. 然而，要求出上述这些特征，需要知道随机过程的一维、二维分布函数，而这些分布函数一般在实际问题中是没有给定的. 为了在实际工作中获得这些特征，用类似处理随机变数时的办法，即统计方法：通过大量的观察或试验，由所获得的数据求出它们的估计值. 以地震勘探为例，为了求得地震波振幅$\{\xi_t\}$在 t_0 时的数学期望 $E(\xi_{t_0})$，我们可在相同条件下放 n 次炮，获得 n 个记录，如图 10.1.1. 把每一记录(一个现实)在 t_0 时的值记录下来，得到 n 个数 $x_1, x_2, \cdots, x_n$，把这 n 个数求平均得$\frac{1}{n}\sum_{i=1}^{n}x_i$，将它作为 $E(\xi_{t_0})$ 的估计值. 诚然，试验次数越多，这个估计值就越接近 $E(\xi_{t_0})$. 但从生产上看，则希望试验次数越少越好，以节省时间，降低生产费用，等等，特别是对于破坏性试验，更是如此. 能否根据一次试验所获得的一个现实来确定随机过程特征呢？这一节将考虑这个问题.

引理 10.7.1 设$\{\xi_t\}$为一均方连续弱平稳过程，其数学期望$E(\xi_t)$为常数 m，则

$$\xi=\frac{1}{2c}\int_{-c}^{c}\xi_t\,dt$$

的数学期望和方差分别为：

$$E(\xi)=E(\xi_t)=m,$$

$$\sigma_\xi^2=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}R(t_1-t_2)\,dt_1dt_2, \qquad (10.7.1)$$

其中 $R(t_1-t_2)=B(t_1-t_2)-m^2$，$B(\tau)$为$\{\xi_t\}$的相关函数。

证 因为

$$E(\xi)=\frac{1}{2c}E\left(\int_{-c}^{c}\xi_t\,dt\right)=\frac{1}{2c}\int_{-c}^{c}E(\xi_t)\,dt=m;$$

$$E(\xi^2)=\frac{1}{4c^2}E\left(\int_{-c}^{c}\xi_{t_1}dt_1\cdot\int_{-c}^{c}\xi_{t_2}dt_2\right)$$

$$=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}B(t_1-t_2)\,dt_1dt_2.$$

故

$$\sigma_\xi^2=E(\xi^2)-[E(\xi)]^2$$

$$=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}B(t_1-t_2)dt_1dt_2-\frac{1}{4c^2}\left[\int_{-c}^{c}E(\xi_t)dt\right]^2$$

$$=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}[B(t_1-t_2)-E(\xi_{t_1})E(\xi_{t_2})]dt_1dt_2$$

$$=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}R(t_1-t_2)dt_1dt_2. \blacksquare$$

(10.7.1)式可以进一步化简，为此令 $t_1-t_2=\tau$，因而有

$$\sigma_\xi^2=\frac{1}{4c^2}\int_{-c}^{c}\int_{-c}^{c}R(t_1-t_2)dt_1dt_2$$

$$=\frac{1}{4c^2}\int_{-2c}^{2c}(2c-|\tau|)R(\tau)d\tau$$

$$=\frac{1}{2c}\int_{-2c}^{2c}\left(1-\frac{|\tau|}{2c}\right)[B(\tau)-m^2]d\tau. \qquad (10.7.2)$$

定理 10.7.1 设$\{\xi_t\}$满足引理 10.7.1 的条件，则 $\underset{c\to\infty}{\text{l.i.m}}\frac{1}{2c}\int_{-c}^{c}\xi_t dt=E(\xi_t)$的充分必要条件为

$$\lim_{c\to\infty}\frac{1}{2c}\int_{-2c}^{2c}\left(1-\frac{|\tau|}{2c}\right)[B(\tau)-m^2]d\tau=0.$$

证 由 (10.7.2) 式及假设立即得到所要证的结果。事实上，这是因为

$$E\left|\frac{1}{2c}\int_{-c}^{c}\xi_t dt-E(\xi_t)\right|^2=\sigma_t^2$$

$$=\frac{1}{2c}\int_{-2c}^{2c}\left(1-\frac{|\tau|}{2c}\right)[B(\tau)-m^2]d\tau$$

之故. ∎

满足定理 10.7.1 条件的随机过程，称为对数学期望具有遍历性，或说是埃尔哥德的.

例 10.7.1 例 10.4.2 的弱平稳过程 $\{\xi_t\}$ 对数学期望是遍历的，事实上，因为其相关函数 $B(\tau)=\cos 2\pi\tau$ 满足定理 10.7.1的条件，即

$$\lim_{c\to\infty}\frac{1}{2c}\int_{-2c}^{2c}\left(1-\frac{|\tau|}{2c}\right)\cos 2\pi\tau d\tau$$

$$=\lim_{c\to\infty}\frac{1}{4c^2}\left[\frac{2}{4\pi^2}(1-\cos 4\pi c)\right]=0.$$

第十一章　概率统计在计算方法中的一些应用

§11.1　蒙特卡罗方法与随机数

一、什么叫蒙特卡罗方法

蒙特卡罗方法又称统计试验方法，它是一种采用统计抽样理论近似地求解数学问题或物理问题的方法．它既可用来研究概率问题，也可用来求解非概率问题．本章的目的只限于介绍一下蒙特卡罗方法的基本思想和原理，而不准备叙述这种方法的全部内容和细节．为使读者较容易了解为什么可以利用概率统计的理论和方法来解决某些数学分析问题，从而抓住蒙特卡罗方法的思想实质，我们着重从求解非概率问题中选取一些较简单而又有启发性的例子加以说明．

利用蒙特卡罗方法去解决数学分析问题时，基本的想法是首先建立与描述该问题有相似性的概率模型．利用这种相似性把这概率模型的某些特征(如随机事件的概率或随机变量的平均值等)与数学分析问题的解答(如积分值、微分方程的解等)联系起来，然后对模型进行随机模拟或统计抽样，再利用所得结果求出这些特征的统计估计值作为原来的分析问题的近似解(如果需要的话还要对解的精度作出估计)．下面我们将要看到，蒙特卡罗方法的理论基础是概率论中很一般的定理——大数定律，因此这方法的应用范围从原则上说来几乎没有什么限制．但是这方法的统计特征又决定它和任何一种统计试验一样，如欲获得充分可靠的结果，则

必须花费大量的劳动和时间，这就使它在实用上受到一定的限制。幸好电子数字计算机的普遍应用，大大地减少了这种限制，因此从本世纪五十年代以来，蒙特卡罗方法得到广泛的发展和应用。特别是随着科学技术的发展，在生产斗争和科学实验过程中，出现了许多复杂难解的问题。例如核物理中描述质点运动的迁移运动方程；大型系统的可靠性分析；地震波的模拟试验；高维数学问题求解；多元统计分析；医学、技术中的诊断、识别；大型生产过程中的运筹规划等等，用传统的物理试验或数学方法进行处理十分困难，这时进行蒙特卡罗模拟常常能够提供一种有效而可行的解决办法。

把随机模拟方法用于近似数值计算领域已有近百年的历史，早期采用这方法的一个著名例子是所谓蒲丰问题(见例1.2.9)：历史上有人作过一次试验，方法是取 $2a=45\text{mm}$，$2l=36\text{mm}$，投针5000次，有2532次针与平行线相交。故

$$p=\frac{2532}{5000}=0.5064,$$

$$\pi=\frac{2l}{ap}\approx\frac{36\times2}{45\times0.5064}=3.160$$

误差约为0.55%。

我们再来看一下利用蒙特卡罗方法计算积分的简单例子。考虑积分

$$I=\int_0^1 f(x)\,dx.$$

假设当 $x\in[0,1]$ 时 $0\leqslant f(x)\leqslant 1$，这时积分 I 等于由曲线 $y=f(x)$，Ox 轴，Oy 轴以及 $x=1$ 围成的区域 G 的面积(见图 11.1.1)。为求此面积，我们设想在正方形 $\{0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1\}$ 内随机地投掷一个点，落点的两个坐标是相互独立且有在区间[0,1]上的均匀分布(即每点都具有等可能性)。那么，这个点落在曲线 $y=f(x)$ 以下的区域 G 内的概率 p 显然等于这区域的面积。以 $(\xi,\ \eta)$ 表示正

方形内任一点的坐标. 如果我们用某种方法得到均匀分布的独立变量 ξ 及 η 的 n 个取样值，对 (ξ,η) 的每一取样值 (ξ_i,η_i) $(i=1,2,\cdots,n)$ 检查 η_i 与 $f(\xi_i)$ 是否满足不等式

$$\eta_i<f(\xi_i) \tag{11.1.1}$$

如果(11.1.1)式成立，说明点 (ξ_i,η_i) 落在区域 G 内[如图 11.1.1 中的点 (ξ_1,η_1)]，否则落在区域 G 外[如图11.1.1中的点 (ξ_2,η_2)]. 设满足不等式(11.1.1)的点数为 m，则由大数定律知当 n 足够大时频率 $\mu=\frac{m}{n}$ 近似于点落在区域 G 的概率 p，亦即 $\frac{m}{n}$ 是所求积分的近似值.

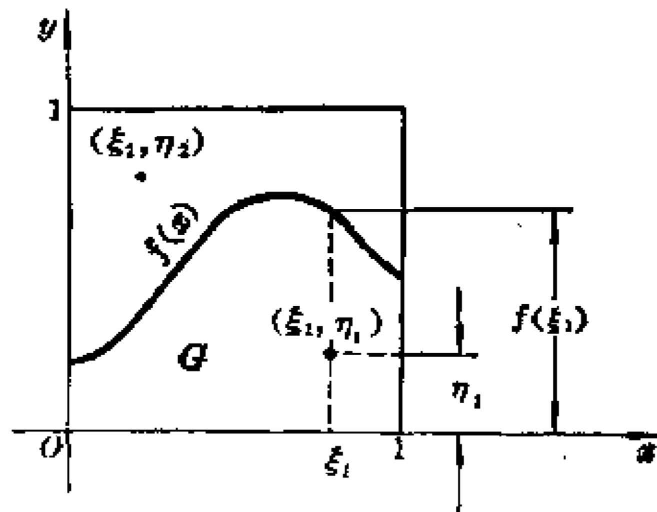


图 11.1.1

二、随机数的产生及伪随机数

在上面的积分计算中，要求获得在(0,1)上均匀分布的独立随机变数 ξ 与 η 的一系列取样值(即随机变量的现实). 一般地说，利用蒙特卡罗方法作各种类型数值计算时，同样也必须找出模拟随机变数或随机过程的现实，为此就需要所谓随机数. 通常实现一次随机变数或随机过程模拟所需的随机数的数量往往是很大的，简单的问题要几千个，复杂的问题则可能要几十万个或更多. 所以，用蒙特卡罗方法计算问题时，绝大部分的计算工作都是有关随机数的运算. 因此，可以这样说，能否有简便、经济、可靠的方法产生

随机数是在实际中判断能否应用蒙特卡罗方法的关键.

最常用的随机数是在区间(0, 1)内均匀分布的随机数(显然，这些随机数可看作是在区间(0, 1)内均匀分布的随机变数的现实)，而且后面我们将要指出任意其它分布律的随机数可以利用均匀分布的随机数来产生，因此我们着重讨论这种随机数. 下面概略地介绍产生随机数的几种常用方法.

(1) 利用专门的随机数表：有一些已制备好的随机数表可供我们计算时选用，例如，本书的附表 7 就是一种在 0 到 99999 的区间内均匀分布的随机数表. 这些表除了可供手算使用外，原则上也可在电子计算机上使用(把随机数表输入计算机储存起来备用). 但是由于计算时常常需要大量的随机数而电子计算机的存储容量又有限，故实用中一般是采用其它方法来获得随机数.

(2) 用随机数发生器产生随机数：考虑取 0 和 1 两个值的随机数序列$\{\xi_i\}$ $(i=1,2,\cdots)$，对每一 i 均有 $P(\xi_i=1)=P(\xi_i=0)=\frac{1}{2}$，例如，随机地抛掷一枚均匀的硬币时，如出现背面令 $\xi_i=1$，出现正面则令 $\xi_i=0$，这就是一种最简单的随机变数发生器. 利用这种发生器可以相等的概率得到 0 和 1，从而就得到一随机变数序列$\{\xi_i\}$. 再利用$\{\xi_i\}$定义如下的二进制数：

$$x=\xi_1\cdot 2^{-1}+\xi_2\cdot 2^{-2}+\cdots+\xi_n\cdot 2^{-n}+\cdots \qquad (11.1.2)$$

可以证明，x 是在区间(0,1)内均匀分布的随机数. 但是在实际计算中数字位数不可能是无限的，我们不得不限制 x 的位数. 例如一台以 k 位二进制符号来表示数的计算机就只能写下 k 位二进制数，总共能写出 2^k 个不同的这种数. 因此，我们实际上只能用有限多个位数是有限的离散随机数(其中每一个数出现的概率都相等，这样得到的分布有时叫做准均匀分布)来代替无限多个均匀分布的连续随机数. 由于目前已有的电子数字计算机其二进制位数

一般都在30位以上，故在利用蒙特卡罗方法时，随机数的离散性对计算的精度影响并不显著.

实际中常用来产生随机数的物理装置有放射粒子计数器、电子管随机数发生器等.

(3) 用专门的运算程序在电子计算机算出的伪随机数：这是指按照一定的计算方法用计算机产生的数. 它们具有类似于均匀分布随机变数的独立取样值的性质. 因为这些数是按照决定性的算法计算出的，所以它们不是真正的随机数，所得数列经过一定时间会出现周期性的重复. 但是，如果计算方法选得恰当，它们就近似地是相互独立和均匀分布的，能够经得起数理统计中的独立性检验和均匀分布检验. 由于以上原因，人们把这种数称做伪随机数，产生伪随机数的最大优点是它的产生和检验都可用电子数字计算机来实现，速度快，而且占用内存小.

三、伪随机数的产生方法

伪随机数通常是利用递推公式产生，方法是定出一个递推公式

$$\xi_n = f(\xi_{n-1}, \xi_{n-2}, \cdots, \xi_{n-k}), \qquad (11.1.3)$$

给定 k 个初始值 ξ_1, ξ_2, …, ξ_k, 我们就可以按上式算出第 $k+1$ 个数

$$\xi_{k+1} = f(\xi_k, \xi_{k-1}, \cdots, \xi_1).$$

一般地，当前面 k 个数 $\xi_{n-1}, \xi_{n-2}, \cdots \xi_{n-k}$ 已算出，下一数 ξ_n 即可按递推公式(11.1.3)算出.

递推公式有各种各样的具体形式，下面介绍其中有代表性的两种.

1° 平方取中法

这是最早产生随机数的一种方法. 一个 b 进制 m 位数 ξ，自乘后一般得到一个 $2m$ 位数 ξ^2. (若不是 $2m$ 位则在前面添加0

补够 $2m$ 位)

$$\xi_0 = e_1 e_2 \cdots e_{m-1} e_m,$$
$$\xi_0^2 = e_1' e_2' \cdots e_{2m-1}' e_{2m}'.$$

取 ξ_0^2 中间的 m 位数(即中间的 m 个数 $e'_{[\frac{m}{2}]+1}, e'_{[\frac{m}{2}]+2}, \cdots, e'_{[\frac{m}{2}]+m}$ ①)

作出如下的 b 进制 m 位数

$$\xi_1 = e'_{[\frac{m}{2}]+1} e'_{[\frac{m}{2}]+2} \cdots e'_{[\frac{m}{2}]+m}.$$

重复以上过程,取 ξ_1^2 的中间 m 位数作 ξ_2,如此下去可得 b 进制 m 位数数列 $\xi_0, \xi_1, \xi_2, \cdots$. 令 $\zeta_n = \xi_n b^{-m}$, 则 $\{\zeta_n\}$ 就是所要求的伪随机数序列.

例11.1.1 考虑 10 进制 4 位数,设 $\xi_0 = 3281$,则

$$\zeta_0 = 3281 \times 10^{-4} = 0.3281,$$
$$\xi_0^2 = 10764961.$$

取中间的 4 位数作 $\xi_1 = 7649$,则

$$\zeta_1 = 7649 \times 10^{-4} = 0.7649.$$
$$\xi_1^2 = 58507201.$$

取中间的 4 位数作 $\xi_2 = 5072$,则

$$\zeta_2 = 5072 \times 10^{-4} = 0.5072.$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

于是得伪随机数序列 0.3281,0.7649,0.5072,…….

例11.1.2 考虑二进制 5 位数,设 $\xi_0 = 01101$,则

$$\zeta_0 = 01101 \times 2^{-5} = \frac{0}{2} + \frac{1}{4} + \frac{1}{8} + \frac{0}{16} + \frac{1}{32} \approx 0.41,$$
$$\xi_0^2 = 0010101001.$$

因为本例中 $m = 5$ 是奇数, $\frac{m}{2} = 2.5$,我们取其整数部分 2,即

① $\left[\frac{m}{2}\right]$ 表示 $\frac{m}{2}$ 的整数部分.

在 ξ_0^2 中取 ε_3^1 到 ε_7^1 共五个数作 ξ_1，于是有 $\xi_1=10101$，则

$$\zeta_1=10101\times 2^{-5}=\frac{1}{2}+\frac{0}{4}+\frac{1}{8}+\frac{0}{16}+\frac{1}{32}\approx 0.66,$$

$$\xi_1^2=0110111001.$$

取中间五个数(第 3 个到第 7 个)作 $\xi_2=10111$，则

$$\zeta_2=10111\times 2^{-5}=\frac{1}{2}+\frac{0}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}\approx 0.72,$$

………………………………………

应用平方取中法时可能遇到"退化"的危险，即出现所有位数的数字均变为零或形成重复循环的序列等现象，同时计算也较复杂. 因此，现在在实际计算中已很少采用这种方法. 但是，先了解这种简单的方法将会有助于我们领会产生伪随机数的基本思想.

2° 同余法

在产生伪随机数的数学方法中，同余法是使用较广的方法，其中又以乘同余法和混合同余法应用最广，因为它们具有能够产生周期长，统计性质好的伪随机数的优点.

乘同余法：用以产生伪随机数的递推公式是

$$x_{n+1}\equiv\lambda x_n(\mathrm{mod}\,M)①, \tag{11.1.4}$$

其中 λ 是乘因子，M 是模数. 当给定了一个初值 x_0 之后，就可以利用上式算出序列 $x_1, x_2, x_3, \cdots, x_n, \cdots$，再取

$$z_n=x_n/M, \tag{11.1.5}$$

即得我们所需要的伪随机数序列.

从上述构造过程可知每一个 x_n（因而 z_n 也一样）最多有 M 个相异值，即 $0\leqslant x_n\leqslant M$（$0\leqslant z_n\leqslant 1$），这表示序列 $\{x_n\}$（$\{z_n\}$）是有周期 L 的，$L\leqslant M$. 因此 $\{z_n\}$ 就不是真正的随机数序列. 但是若 L 充

① 这式子称做以 M 为模数的同余式（mod是modulus 的缩写），即是以 M 除 λx_n 后得到的余数记为 x_{n+1}.

分大，则在同一个周期内的数有可能经受得住数理统计中的独立性和均匀性检验，这完全取决于参数 x_0, λ, M 的选择. 至于如何选取参数，主要依靠通过计算机作试验. 有一些文献报导如下一些参数组，可供使用时参考.

$$x_0=1, \lambda=7, M=10^{10} \quad (L=5\times 10^7);$$

$$x_0=1, \lambda=5^{13}, M=2^{36} \quad (L=2^{34}\approx 2\times 10^{10});$$

$$x_0=1, \lambda=5^{17}, M=2^{42} \quad (L=2^{40}\approx 10^{12})$$

(在二进制数字计算机中，如尾部字长为 k，一般取 $M=2^k$).

例11.1.3 取参数组 $x_0=1, \lambda=7, M=10^3$ 则

$$\lambda x_0=7\times 1=7, \quad \therefore \quad x_1=7, \quad z_1=\frac{7}{1000}=0.007.$$

$$\lambda x_1=7\times 7=49, \quad \therefore \quad x_2=49, z_2=\frac{49}{1000}=0.049,$$

$$\lambda x_2=7\times 49=343, \therefore \quad x_3=343, \quad z_3=\frac{343}{1000}=0.343,$$

$$\lambda x_3=7\times 343=2401 \quad \therefore \quad x_4=401, \quad z_4=\frac{401}{1000}=0.401,$$

$$\lambda x_4=7\times 401=2807 \quad \therefore \quad x_5=807, z_5=\frac{807}{1000}=0.807,$$

…如此继续下去.

混合同余法：递推公式是

$$x_{n+1}\equiv \lambda x_n+C(\mathrm{mod} M), \tag{11.1.6}$$

其中 C 是非负整数. 通过适当选取参数 C 可以改善伪随机数的统计性质.

组合同余法：实际应用和统计检验表明，用上述两种同余法产生的伪随机序列，在高维情况下会出现许多在统计上不能接受的缺陷，需要作进一步的改善，组合同余法就是其中最简单的一种，其具体作法如下：

先用乘同余法的递推公式(11.1.4)产生伪随机数列$\{z_n\}$. 又用混合同余法的递推公式(11.1.6)(或其它合适的方法)产生另一个伪随机数z^*. 然后利用z^*的前面若干位数字在$z_1, z_2, \cdots, z_p$(即数列$\{z_n\}$中的前p个数)中随机地选取一个作为实际使用的伪随机数，这里p是一个预先适当确定的正整数. 设这样选得的伪随机数是$z_{n_1} = z_l (1 \leqslant l \leqslant p)$，再用同样的方法在$z_{l+1}, \cdots, z_{l+p}$这$p$个数中选取第二个实际使用的伪随机数$z_{n_2}, \cdots$. 如此继续下去就可以从数列$\{z_n\}$中选出一个子列$\{z_{n_k}\}$，这就是实际使用的伪随机数序列.

例如，在二进制数字计算机的实际计算中，可以取$p = 2^5 = 32$. 令

$$\lambda_l \equiv \lambda^l \pmod{M}, \quad l = 1, 2, \cdots, 32,$$

这里的λ是(11.1.4)式中的乘子. 用混合同余法产生一个(非规格化的)伪随机数z^*，用z^*的前面五位二进制数字确定l(从而也就确定了λ_l)，然后用

$$x_{n_{k+1}} \equiv \lambda_l x_{n_k} \pmod{M} \qquad (11.1.7)$$

代替(11.1.4)，即可产生实际使用的伪随机数列$\{z_{n_k}\}$. 应当指出，我们在这里并没有先如实地产生数列$\{z_n\}$，而是直接算出它的子列$\{z_{n_k}\}$，这就大大地减少了计算工作量.

最后，简单地提一下所谓准随机数，设θ是一无理数，令

$$x_n = \{n\theta\} \quad (n = 1, 2, 3, \cdots),$$

这里$\{x\}$表示实数x的小数部分. 由分析理论知x_n在(0,1)上处处稠密地均匀分布着，在求解的实际问题中，若对伪随机数的要求不高(如用蒙特卡罗方法计算高维数值积分时)，则可用准随机数x_n代替伪随机数进行计算而获得较快的收敛速度.

四、随机数的统计检验

不管用哪一种方法产生的随机数序列都存在这样的问题，即

能否在实际中把它们看作是在(0，1)上均匀分布的连续随机变数的独立取样值．因此，我们必须对它们进行统计检验，看看它们是否具有较好的独立性(即随机性)和均匀性，然后决定能否在蒙特卡罗方法中使用．下面介绍几种简单的检验方法．但是应该指出，统计检验不能完全肯定或否定某一假设，因此最好能多作几种统计检验，以便有较大的把握保证使用的随机序列有较好的统计性质．

① 参数检验

在这里介绍的参数检验是子样平均值和理论平均值的差异显著性检验

在(0，1)上均匀分布的随机变数 ξ 和 ξ^2 的平均值及方差分别为

$$E(\xi)=\int_0^1 x\,dx=\frac{1}{2};$$

$$E(\xi^2)=\int_0^1 x^2\,dx=\frac{1}{3};$$

$$D(\xi)=E(\xi^2)-[E(\xi)]^2=\frac{1}{3}-\frac{1}{4}=\frac{1}{12};$$

$$D(\xi^2)=E(\xi^4)-[E(\xi^2)]^2=\int_0^1 x^4\,dx-\left(\frac{1}{3}\right)^2$$

$$=\frac{1}{5}-\frac{1}{9}=\frac{4}{45}.$$

如果 N 个随机数 $z_1, z_2, \cdots, z_N$ 是 ξ 的 N 个独立观测值，令

$$\bar{z}=\frac{1}{N}\sum_{i=1}^{N} z_i, \qquad \overline{z^2}=\frac{1}{N}\sum_{i=1}^{N} z_i^2,$$

则它们的平均值和方差为:

$$E(\bar{z})=\frac{1}{2};$$

$$E(\overline{z^2})=\frac{1}{3};$$

$$D(\bar{z})=\frac{1}{N^2}\times ND(\xi)=\frac{1}{12N};$$

$$D(\overline{z^2})=\frac{1}{N^2}\times ND(\xi^2)=\frac{4}{45N}.$$

又由中心极限定理(参看§5.4)知统计量

$$u_1=\frac{\bar{z}-E(\bar{z})}{\sqrt{D(\bar{z})}}=\sqrt{12N}\left(\bar{z}-\frac{1}{2}\right), \qquad (11.1.8)$$

$$u_2=\frac{\overline{z^2}-E(\overline{z^2})}{\sqrt{D(\overline{z^2})}}=\frac{1}{2}\sqrt{45N}\left(\overline{z^2}-\frac{1}{3}\right) \qquad (11.1.9)$$

渐近地服从正态 $N(0,1)$，故当给定显著性水平后，即可根据正态分布表确定临界值，据此判断 $\bar{z}$ 与 ξ 的平均值 $E(\xi)$ 和 $\overline{z^2}$ 与 ξ^2 的平均值 $E(\xi^2)$ 之差异是否显著，从而决定能否把 $z_1,\cdots,z_N$ 看作是(0,1)上均匀分布的随机变数 ξ 的 N 个独立取样值.

② 分布均匀性检验

分布均匀性检验又称频率检验，它用来检验经验频率和理论频率是否有显著差异.

把(0,1)区间分成 k 等分，以 $\left(\frac{i-1}{k},\frac{i}{k}\right)(i=1,2,\cdots,k)$ 表示第 i 个小区间. 如 z_s 是(0,1)上均匀分布的随机变数 ξ 的一个取样值，则它落在任一小区间的概率 p_i 皆应等于这些小区间的长度 $\frac{1}{k}$，故 N 个 z_s 值落在任一个小区间的平均数 $m_i=Np_i=\frac{N}{k}$. 设实际上 $z_1,z_2,\cdots,z_N$ 中属于第 i 个小区间的数目为 n_i，则统计量

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-m_i)}{m_i}=\frac{k}{N}\sum_{i=1}^{k}\left(n_i-\frac{N}{k}\right) \qquad (11.1.10)$$

渐近地服从自由度为 $(k-1)$ 的 χ^2-分布(参看§8.3).

我们还可以作累积频率检验，这时利用统计量

$$\sqrt{N}D_N=\frac{1}{\sqrt{N}}\max_{1\leqslant s\leqslant k}\left|\sum_{i=1}^{s}\left(n_i-\frac{N}{k}\right)\right| \qquad (11.1.11)$$

渐近地服从柯尔莫哥洛夫-斯米尔诺夫分布这一事实进行显著性检验.

③ 独立性检验

独立性检验主要是检验随机数 $z_1,\cdots,z_N$ 中前后各数的统计相关性是否显著，下面介绍两种检验.

1° 相关系数检验

两个随机变数的相关系数反映它们之间的线性相关程度. 若两个随机变数相互独立，则它们的相关系数必为零(反之不一定)，故可以利用相关系数来检验随机数的独立性.

设给定 N 个随机数 $z_1,z_2,\cdots,z_N$，我们计算前后距离为 k 的样本相关系数

$$r_k=\left[\frac{1}{N-k}\sum_{i=1}^{N-k}z_iz_{k+i}-(\bar{z})^2\right]\Big/s^2 \quad (k=1,2,\cdots), \qquad (11.1.12)$$

其中 $s^2=\frac{1}{N}\sum_{i=1}^{N}(z_i-\bar{z})^2$. 如各 z_i 相互独立则相关系数应为零，在原假设 $r_k=0$ 之下，当 N 充分大(例如 $N-k>50$)时统计量

$$r_k\sqrt{N-k} \qquad (11.1.13)$$

渐近地服从正态 $N(0,1)$. 具体应用时可对若干个不同的 k 值作检验.

2° 联列表检验

在 xy 平面上，将单位正方形分为 k^2 个相等的小正方形，又把 N 个随机数 $z_1,z_2,\cdots,z_N$ 按出现的先后顺序两两分组，例如可取

$$(z_1,z_{l+1}),(z_2,z_{l+2}),\cdots,(z_{N-l},z_N),(z_{N-l+1},z_1),\cdots,(z_N,z_l).$$

记这些数对落在第 (i, k) 个小正方形内的数目为 $n_{ik}(i, k=1, 2, \cdots, N)$. 令

$$n_{i\cdot}=\sum_{k=1}^{N} n_{ik}, \qquad n_{\cdot k}=\sum_{i=1}^{N} n_{ik}.$$

根据独立性假设,统计量

$$\chi^2=N\left(\sum_{i,k=1}^{N} \frac{n_{ik}}{n_{i\cdot}n_{\cdot k}}-1\right) \tag{11.1.14}$$

渐近地服从自由度为 $(k-1)^2$ 的 χ^2-分布.

随机数的统计检验除了参数检验、均匀性检验和独立性检验之外还有组合规律性检验，连检验等，我们在此就不一一介绍了.读者若要进一步学习可参看文献[34].

§11.2 任意随机变数的模拟

在利用蒙特卡罗方法进行计算而建立的概率模型中，常常会遇到具有各种不同分布的随机变数(一维或多维)，这时就要求产生对应于该随机变数(或分布)的随机数，我们把这一步骤称做对这随机变数进行模拟或抽样.

一、离散型情形

先考虑事件序列的模拟，设事件序列 $A_0, A_1, \cdots, A_i, \cdots$ 满足如下完备性条件

$$A_i \cap A_k=\varnothing \qquad (i \neq k);$$

$$P\left(\bigcup_{i=0}^{\infty} A_i\right)=1.$$

这表示每次试验或观测的结果必有且只有其中的一个事件发生.

令 $p_i=P(A_i)$，$p^{(n)}=\sum_{i=0}^{n}p_i$，于是 $\{p^{(n)}\}$ 将 $(0,1)$ 分成可列多个小区间．现在任取一个 $(0,1)$ 上均匀分布的随机数 r，如果 r 落在第 n 个小区间，即 $p^{(n-1)}<r\leqslant p^{(n)}$，则说事件 A_n 发生（令 $p^{(-1)}=0$）．容易看出 $P\{p^{(n-1)}<r\leqslant p^{(n)}\}=p^{(n)}-p^{(n-1)}=p_n=P(A_n)$，所以这样的模拟合理．

现在讨论离散随机变数 ξ 的模拟．设它取值 x_i 的概率等于 p_i，即 $P\{\xi=x_i\}=p_i(i=0,1,2,\cdots)$．显然事件序列 $A_i=\{\xi=x_i\}$ $(i=0,1,2,\cdots)$ 满足完备性条件，因此我们就可以直接利用上述模拟完备事件序列 $A_0,A_1,A_2,\cdots$ 的方法来模拟 ξ．

在具体计算问题中，p_n 不一定能有简单的表达式，这时利用具体分布本身的某些特征往往能得到简单的模拟方法．

例11.2.1 贝努里概型二项分布的模拟

在§1.5 中我们已经介绍过贝努里概型．这是概率统计中一种最简单而又常用的概型．它由一系列试验组成，其中每一次试验只有两种可能结果——成功与失败，我们分别用事件 A 和 $\bar{A}$ 表示．若事件 A 发生的概率 $P\{A\}=p(0<p<1)$，事件 $\bar{A}$ 发生的概率 $P\{\bar{A}\}=q=1-p$．而且各次试验是相互独立的．易见，A 与 $\bar{A}$ 组成一完备事件系．故可用上述模拟一完备事件序列的方法对贝努里概型进行模拟，即任取一个在 $(0,1)$ 上均匀分布的随机数 r，如果 $r\leqslant p$ 就认为事件 A 发生，反之，若 $r>p$ 则认为事件 $\bar{A}$ 发生．

从贝努里概型出发，结合不同的条件就可以得到许多不同类型的离散随机数，其中常用的有二项分布，负二项分布和几何分布等，下面以二项分布为例谈谈对这些变数的模拟．

在由 n 次独立试验组成的贝努里概型中，事件 A 发生的次数 η 是一随机变数，它取值 $k(k=0,1,\cdots,n)$ 的概率是

$$p_k=P(\eta=k)=C_n^k p^k q^{n-k}. \qquad (11.2.1)$$

对具有二项分布的随机变数的模拟，可以直接根据(11.2.1)式用上述一般方法进行模拟，但当p较大而计算精度又要求较高时，我们也可以在计算机上用n次贝努里试验产生二项分布的随机数，即首先产生n个均匀分布随机数$r_1, r_2, \cdots, r_n$. 然后统计其中使得$r_i \geqslant p (i=1,2,\cdots,n)$的个数$n_i$，这就是我们所要求的随机数.

例11.2.2 求具有参数λ的泊松分布

$$P(k,\lambda)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad (11.2.2)$$

的随机数.

我们可以利用泊松分布的某些特征进行计算.

1° 由泊松定理知道，若进行n次独立试验，在每一次试验中事件A发生的概率等于p_n，则在n次试验中事件A发生k次的概率当$n\to\infty$, $p_n\to 0$，且$np_n=\lambda$时趋于(11.2.2)给出的$P(k,\lambda)$. 基于这一性质，我们可以选取足够大的n，使得$p_n=\frac{\lambda}{n}$相当小（例如不大于0.1). 然后作n次独立试验，使每一试验中事件A发生的概率等于p_n. 如以η表示在n次试验的序列中事件A实际上发生的次数，则η近似地服从泊松分布，而且n愈大愈准确.

2° 由(11.2.2)式直接看出$P(0,\lambda)=e^{-\lambda}$，而且我们还可证明$n$个相互独立、在$(0,1)$上均匀分布的随机变数$\xi_1, \xi_2, \cdots, \xi_n$的乘积$\xi_1, \xi_2, \cdots, \xi_n$的分布密度$f(x)=\frac{(-\ln x)^{n-1}}{(n-1)!}$（参看[5]）. 基于这些事实，我们取一串均匀分布随机数$\{r_i\}$；若$r_1<e^{-\lambda}$，令$\eta=0$；若$r_1\geqslant e^{-\lambda}$，则找出满足不等式

$$r_1\cdot r_2\cdots r_k \geqslant e^{-\lambda} > r_1\cdot r_2\cdots r_k\cdot r_{k+1} \qquad (11.2.3)$$

的整数k（即第一次使得$e^{-\lambda}>r_1\cdot r_2\cdots r_k\cdot r_{k+1}$成立的$k$）并令$\eta=k$. 因为

$$P\{\eta=k\}=P\{r_1\cdot r_2\cdots r_k\geqslant e^{-\lambda}>r_1\cdot r_2\cdots r_k\cdot r_{k+1}\}$$
$$=P\{r_1\cdot r_2\cdots r_k\geqslant e^{-\lambda}\}-P\{r_1\cdot r_2\cdots r_k\cdot r_{k+1}\geqslant e^{-\lambda}\}$$
$$=\int_{e^{-\lambda}}^{1}\frac{(-\ln x)^{k-1}}{(k-1)!}dx-\int_{e^{-\lambda}}^{1}\frac{(-\ln x)^{k}}{k!}dx,$$

用分部积分法算出上面的积分等于

$$-x\frac{(-\ln x)^k}{k!}\bigg|_{e^{-\lambda}}^{1}=e^{-\lambda}\cdot\frac{(\lambda)^k}{k!}$$

这就证明了 η 是具有泊松分布(11. 2. 2)的随机变数.

二、一维连续型情形

一般说来，具有给定分布的一维连续随机变数可以利用在(0,1)上均匀分布的随机数来模拟，这就是说具有给定分布的随机数可以由在(0,1)上均匀分布的随机数通过某些变换而产生. 由此可见上节所讨论的均匀分布随机数是进行统计模拟的基础. 现在介绍几种常用的模拟方法.

(一) 反函数法

本法利用如下的已知结果：若随机变数 η 具有连续分布函数 $F(x)$，而 ξ 是在(0,1)上均匀分布的随机变数，它的分布函数记为 $G(x)$①，则

$$\eta=F^{-1}(\xi). \tag{11. 2. 4}$$

事实上，我们有

$$P(F^{-1}(\xi)<x)=P(\xi<F(x))=G(F(x))=F(x).$$

这就证明了 $F^{-1}(\xi)$的分布函数为 $F(x)$. (11. 2. 4)式表示 $F^{-1}(\xi)$ 是方程

$$\xi=F(\eta) \tag{11. 2. 5}$$

---

① 回忆

$$G(x)=\begin{cases}0, & 若\ x\leqslant 0;\\ x, & 若\ x\in(0,1);\\ 1, & x\geqslant 1.\end{cases}$$

的解. 若进一步假设 η 有分布密度 $f(x)$,则(11.2.5)可写成

$$\xi=\int_{-\infty}^{\eta} f(x)\,dx. \tag{11.2.6}$$

因此,如果给出了在(0,1)上均匀分布的随机数 r_i,则具有给定分布的随机数 y_i 可由求解方程

$$r_i=\int_{-\infty}^{y_i} f(x)\,dx \tag{11.2.7}$$

得出.

例11.2.3 对任意 $a<b$,求在(a,b)上均匀分布的随机数.

已知在(a,b)上均匀分布的随机变量有密度函数

$$f(x)=\begin{cases}0 & ,\ 若\ x\leqslant a;\\ (b-a)^{-1}, & 若\ x\in(a,b);\\ 0 & ,\ 若\ x\geqslant b.\end{cases}$$

故由(11.2.7)式有

$$r_i=\int_a^{y_i}\frac{dx}{b-a}=\frac{y_i-a}{b-a}.$$

由此得

$$y_i=a+(b-a)r_i. \tag{11.2.8}$$

例11.2.4 求具有指数分布密度函数 $f(x)=\lambda e^{-\lambda x}(x\geqslant 0)$的随机数 y_i,其中 λ 是指数分布的参数.

由(11.2.7)式有

$$r_i=\int_0^{y_i}\lambda e^{-\lambda x}dx=1-e^{-\lambda y_i}.$$

由此得

$$y_i=-\frac{1}{\lambda}\ln(1-r_i).$$

因为 $1-r_i$ 与 r_i 都是 (0, 1) 上均匀分布的随机数,故上式可简化为

$$y_i=-\frac{1}{\lambda}\ln r_i. \tag{11.2.9}$$

这种变换方法是一种普遍方法，但当反函数不存在或难以求出，例如，在不能由积分式(11.2.7)得到 y_i 的显式表示时，我们只能将分布函数表输入电子计算机储存起来备用。这时反函数法就不宜于使用，因此需要改用不必求解(11.2.7)式的模拟方法。下面介绍几种这样的方法。

（二）舍选法

这方法的实质是从许多均匀随机数中选出一部分，使之成为具有给定分布的随机数。

我们要模拟具有给定密度函数 $f(x)$ 的随机变数，设 $f(x)$ 集中于有限区间 (a,b) $\left(即\int_a^b f(x)\,dx=1\right)$ 且在 (a,b) 上有界。于是可选取正常数 α，使得 $\alpha f(x)<1$，对一切 $x\in(a,b)$ 成立。又设 ξ、η 是两个独立的随机变数，ξ 在 (a,b) 上均匀分布，而 η 在 $(0,1)$ 上均匀分布。于是可以证明

$$P(\xi\leqslant d\,|\,\alpha f(\xi)\geqslant\eta)=\int_a^d f(x)\,dx\quad(a<d<b).\qquad(11.2.10)$$

这就是说在条件 $\alpha f(\xi)\geqslant\eta$ 之下，ξ 的条件分布密度为 $f(x)$。

事实上，由假设得 (ξ,η) 的密度函数是

$$g(x,y)=\begin{cases}\dfrac{1}{b-a}, & x\in(a,b),\ y\in(0,1),\\ 0, & 其它情形.\end{cases}$$

故

$$P(\xi\leqslant d\,|\,\alpha f(\xi)\geqslant\eta)=\iint\limits_{\substack{\alpha f(x)\geqslant y\\ a<x\leqslant d}}\frac{dx\,dy}{b-a}\Bigg/\iint\limits_{\alpha f(x)\geqslant y}\frac{dx\,dy}{b-a}$$

$$=\int_a^d\int_0^{\alpha f(x)}dy\,dx\Bigg/\int_a^b\int_0^{\alpha f(x)}dy\,dx$$

$$=\int_a^d\alpha f(x)\,dx\Bigg/\int_a^b\alpha f(x)\,dx=\int_a^d f(x)\,dx.$$

根据上述，我们可以先取一列在(a, b)上均匀分布的随机数$\xi_1, \xi_2, \cdots, \xi_n$，再取一列在$(0,1)$上均匀分布的随机数$\eta_1, \eta_2, \cdots, \eta_n$。如果$\alpha f(\xi_1) \geqslant \eta_1$，就取$\xi_1$作为我们所求的第一个$f(x)$分布的随机数，如$\alpha f(x) < \eta_1$，则舍弃之，然后同样考察$\xi_2, \eta_2$，如此继续下去.

此法的直观意义如下：在边长为1和$(b-a)$的矩形$abb'a'$内掷点P，若随机点P位于曲线$\alpha f(x)$的下面（如图11.2.1中的P_1）则以该点的横坐标ξ作为所求随机变数的一个取样值，否则（如图11.2.1中的P_2）舍弃之. 这时在条件$\alpha f(\xi) \geqslant \eta$之下$\{\xi \leqslant d\}$的概率是图中重阴影部分与阴影部分面积之比，即

$$\int_a^d \alpha f(x)\,dx \Big/ \int_a^b \alpha f(x)\,dx = \int_a^d f(x)\,dx.$$

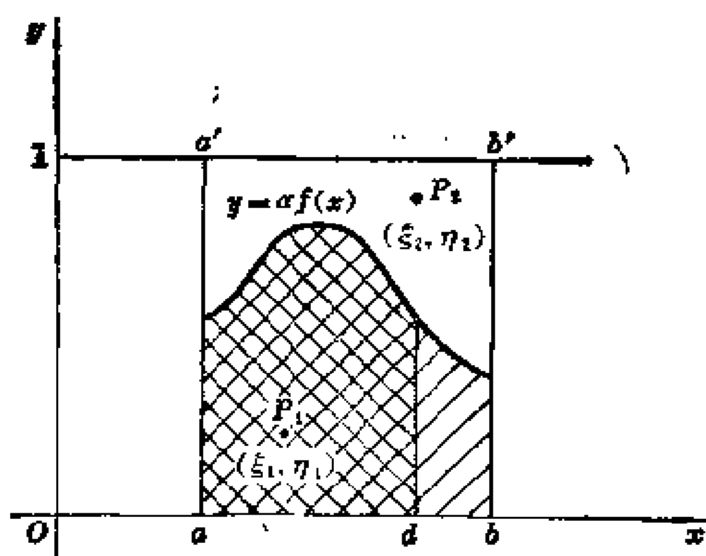


图 11.2.1

上面假定(a, b)是有限区间，如果它不是有限，我们总可以选取有限区间(a, b)，使得

$$\int_{a_1}^{b_1} f(x)\,dx \geqslant 1-\varepsilon,$$

其中ε是任意小的正数，然后对(a_1, b_1)利用上述方法，这时会出现小的系统误差.

这方法适用于 $f(x)$ 变化不激烈的情形. 若 $f(x)$ 在小区间内取值变化很大,这方法就不稳定. 另外,这方法需要计算 $f(\xi)$,因此若这函数表达式很复杂时,计算量也就很大.

(三) 离散近似法

这方法是把连续分布分段集中而作离散化近似. 设密度函数 $f(x)$ 集中在有限区间 (a,b) 上, 将 (a,b) 分为 n 分, 它的分点为 $a=a_0<a_1<\cdots<a_n=b$. 令

$$p_i=\int_{a_{i-1}}^{a_i} f(x)dx,$$

显然有

$$\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}\int_{a_{i-1}}^{a_i} f(x)dx=\int_a^b f(x)dx=1.$$

又将 $(0,1)$ 也分成 n 分, 其分点为 $0=c_0<c_1<\cdots<c_n=1$, 而且有 $c_i-c_{i-1}=p_i$,现在任取均匀随机变数 η, 若 $c_{i-1}<\eta\leqslant c_i$ 则令

$$\xi=a_{i-1}+(a_i-a_{i-1})\frac{\eta-c_{i-1}}{c_i-c_{i-1}}. \qquad (11.2.11)$$

当 n 充分大时, 可以认为 ξ 近似地是 $f(x)$ 分布的随机变数. 事实上

$$P(a_{i-1}<\xi\leqslant a_i)=P(c_{i-1}<\eta\leqslant c_i)=c_i-c_{i-1}$$
$$=p_i=\int_{a_{i-1}}^{a_i} f(x)dx.$$

和舍选法一样,若 (a,b) 不是有限的,则可用(11.2.10)式作截尾方法处理.

(四) 其它方法

除了上述三种之外,还有好些其它模拟方法,在此只是结合一些具体分布及其特性举例介绍其中的一、两种. 读者在实际工作中可以根据具体情况灵活运用或提出新的方法.

例11.2.5 求正态分布$N(a,\sigma)$的随机数.

除了可以利用上述方法之外再介绍另外两种方法

(1) 利用中心极限定理

设$\xi_1,\xi_2,\cdots,\xi_n$是n个相互独立的在(0,1)上均匀分布的随机变数，因为这时有$E(\xi_i)=\frac{1}{2}$，$D(\xi_i)=\frac{1}{12}$，故由中心极限定理(§5.4)知随机变数

$$\eta=\left(\sum_{i=1}^{n}\xi_i-\frac{n}{2}\right)\Big/\sqrt{\frac{n}{12}} \tag{11.2.12}$$

渐近地有正态分布$N(0,1)$. 因此，我们若选取n个在(0,1)上均匀分布的随机数$r_1,r_2,\cdots,r_n$，则

$$x=\left(\sum_{i=1}^{n}r_i-\frac{n}{2}\right)\Big/\sqrt{\frac{n}{12}} \tag{11.2.13}$$

近似地有正态分布$N(0,1)$. 为了保证一定的精确度，上式中的n应取得足够大(一般情况下约取$n=10$即可). 又从(11.2.13)式看出，当$n=12$时分母变为1，这时有最简单的形式.

有了正态$N(0,1)$的随机数x，则由公式

$$y=\sigma x+a \tag{11.2.14}$$

可算出正态$N(a,\sigma)$的随机数y.

(2) 坐标变换法

设ξ_1, ξ_2是两个相互独立的在(0,1)上均匀分布的随机变数，作变换

$$\left.\begin{aligned}\eta_1&=(-2\ln\xi_1)^{\frac{1}{2}}\cos(2\pi\xi_2),\\ \eta_2&=(-2\ln\xi_1)^{\frac{1}{2}}\sin(2\pi\xi_2),\end{aligned}\right\} \tag{11.2.15}$$

则η_1,η_2是两个相互独立的正态$N(0,1)$随机变数. 现在证明这一点. 因为ξ_1,ξ_2的联合密度函数是

$$f(\xi_1,\xi_2)=\begin{cases}1, & \xi_1\in(0,1),\xi_2\in(0,1);\\ 0, & \text{其它情形.}\end{cases}$$

从(11.2.15)解出

$$\left.\begin{aligned}\xi_1&=e^{-\frac{\eta_1^2+\eta_2^2}{2}}\\ \xi_2&=\frac{1}{2\pi}\operatorname{tg}^{-1}\frac{\eta_2}{\eta_1},\end{aligned}\right\} \tag{11.2.16}$$

变换的雅可比行列式是

$$\frac{\partial(\xi_1,\xi_2)}{\partial(\eta_1,\eta_2)}=\begin{vmatrix}-\eta_1 e^{-\frac{\eta_1^2+\eta_2^2}{2}} & -\eta_2 e^{-\frac{\eta_1^2+\eta_2^2}{2}}\\ \frac{1}{2\pi}\frac{-\frac{\eta_2}{\eta_1^2}}{1+\left(\frac{\eta_2}{\eta_1}\right)^2} & \frac{1}{2\pi}\frac{\frac{1}{\eta_1}}{1+\left(\frac{\eta_2}{\eta_1}\right)^2}\end{vmatrix}$$

$$=-\frac{1}{2\pi}e^{-\frac{\eta_1^2+\eta_2^2}{2}}=-\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta_1^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta_2^2}{2}}\right).$$

故 η_1, η_2 的联合分布密度函数是$\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta_1^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{\eta_2^2}{2}}$，即是两个正态$N(0,1)$的密度函数的乘积，故 η_1,η_2 是两个相互独立的正态$N(0,1)$随机变数.

根据(11.2.15)即可通过均匀随机数求出正态$N(0,1)$的随机数，从而由(11.2.14)式便得到正态$N(a,\sigma)$的随机数.

例11.2.6 求具有瑞利分布(即密度函数为

$$f(x)=\begin{cases}\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}, & x>0,\\ 0 & , x\leqslant 0,\end{cases} \tag{11.2.17}$$

其中 $\sigma>0$ 是分布的参数)的随机数.

因为瑞利分布是具有密度函数 $f(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$的二维正态随机变数$(\xi,\eta)$的模 $r=\sqrt{\xi^2+\eta^2}$的分布（参看第二章例

2.4.6) 故可按下述步骤获得具有瑞利分布的随机数 $r_i(i=1,2,\cdots)$.

(1) 求出具有正态$N(0,\sigma)$的随机数 x_i 和 $y_i(i=1,2,\cdots)$.

(2) 算出$r_i=\sqrt{x_i^2+y_i^2}$,这就是参数为σ的瑞利分布随机数.

三、多维连续型情形

现在简单地讨论一下 n 维随机变数（或称随机向量）的模拟，上面就一维情形所介绍的舍选法，离散近似法等都可以应用于多维情形. 例如:在利用舍选法时，假设给定的 n 维分布密度函数. $f(x_1,x_2,\cdots,x_n)$集中于有限的 n 维长方体$[a_1,b_1]\times[a_2,b_2]\times\cdots\times[a_n,b_n]$,且有 $\alpha f(x_1,x_2,\cdots,x_n)<1$ 这时可先选取 $n+1$ 个在$(0,1)$上均匀分布的随机变数 $\zeta_1,\cdots,\zeta_n$ 和 η, 令 $\xi_i=(b_i-a_i)\zeta_i+a_i(i=1,2,\cdots,n)$, 则当

$$\alpha f(\xi_1,\xi_2,\cdots,\xi_n)\geqslant\eta \tag{11.2.18}$$

时，$(\xi_1,\xi_2\cdots,\xi_n)$ 即为我们所求的 n 维 $f(x_1,x_2,\cdots,x_n)$分布随机数. 若(11.2.18)式不成立,则将这组随机数舍弃，另外再产生新的随机数. 如此按上述决定取舍,将过程继续下去.

利用离散近似法也和一维情形完全一样，但当 $f(x_1,x_2,\cdots,x_n)$变化较慢而又容许一定的误差时，n 维点$(\xi_1,\xi_2,\cdots,\xi_n)$落在小区域$\Delta V$ 内的概率可以不必通过积分而用 $f(x_1^0,x_2^0,\cdots,x_n^0)|\Delta V|$来近似,这里$(x_1^0,x_2^0,\cdots,x_n^0)$是$\Delta V$ 中某一点，$|\Delta V|$是 ΔV 的 n 维体积.

下面再介绍所谓条件密度法，这种方法的实质是利用条件分布密度把多维模拟化为一维模拟. 它的优点是适用于一般情形,但当 n 太大和密度函数太复杂时难于实现.

根据联合分布密度函数的性质，对于任意 n 维联合分布密度函数 $f_n(x_1,x_2,\cdots,x_n)$，将其对其中一个变元在全变域积分，即得余下的 $n-1$ 个变元的 $n-1$ 维联合分布密度函数,即

$$\int f_n(x_1,x_2,\cdots,x_n)dx_n = f_{n-1}(x_1,x_2,\cdots,x_{n-1}). \qquad (11.2.19)$$

而根据定义，在条件 $\xi_1=x_1,\xi_2=x_2,\cdots,\xi_{n-1}=x_{n-1}$ 之下 ξ_n 的条件分布密度函数

$$f(x_n|x_1,x_2,\cdots,x_{n-1})=\frac{f_n(x_1,x_2,\cdots,x_n)}{f_{n-1}(x_1,x_2,\cdots,x_{n-1})}. \qquad (11.2.20)$$

反复运用(11.2.20)可得递推公式

$$f_n(x_1,x_2,\cdots,x_n)=f(x_n|x_1,x_2,\cdots,x_{n-1})f(x_{n-1}|x_1,x_2,\cdots,x_{n-2})\cdots f(x_2|x_1)f_1(x_1). \qquad (11.2.21)$$

由此式可得如下模拟方法：利用对一维情形所介绍的方法，先选出具有一维分布密度函数 $f_1(x_1)$ 的随机数 $x_1^{(0)}$，以 $x_1^{(0)}$ 代入 $f(x_2|x_1)$ 中的 x_1，于是得到一个确定的一维分布密度 $f(x_2|x_1^{(0)})$，再选出具有一维分布密度 $f(x_2|x_1^{(0)})$ 的随机数 $x_2^{(0)}$. 一般地，如 $x_1^{(0)}$，$x_2^{(0)}$，…，$x_{k-1}^{(0)}$ 已确定，则可选出具有一维分布密度 $f(x_k|x_1^{(0)},x_2^{(0)},\cdots,x_{k-1}^{(0)})$ 的随机数 $x_k^{(0)}$，照此方式进行下去，最后得到的 n 维向量 $(x_1^{(0)},x_2^{(0)},\cdots,x_n^{(0)})$ 就是 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的模拟.

对于具体的多维分布也和一维情形一样，可以根据分布的不同特点提出各种各样的模拟方法，在此就不一一介绍了.

四、随机游动和马尔可夫链的模拟

因为在后面讨论用蒙特卡罗方法解偏微分方程时需要模拟质点的随机游动，所以我们要介绍一下这类特殊随机过程的模拟方法. 但是应当指出，在许多自然现象和科学实验，例如有关分子运动和原子反应堆的问题中，都要考虑大量质点的随机游动. 一维(直线上的)随机游动是最简单的一种游动模型. 一质点从直线上的某点出发，每次以概率 p 左移一步，以概率 $q=1-p$ 右移一步，质点到达新的位置后，继续以同样的规律移动，直到碰到某边界点而停止游动，这样的边界点称做吸收壁，这种随机游动的模拟方法

是：取在(0,1)上均匀分布的随机数 r_1，如 $r_1<p$，定义 $\xi_1=1$ 表示质点左移一步，如 $r_1\geqslant p$，定义 $\xi_1=-1$，表示质点右移一步，依次取均匀随机数 $r_1,r_2,\cdots,r_n,\cdots$，并分别与 p 比较后得一列 ξ_1，ξ_2，$\cdots,\xi_n,\cdots$

令 $S_n=\sum\limits_{i=1}^{n}\xi_n$，则和数 S_n 表示经 n 步游动后距出发点的步数(正号表示在出发点左边，负号表示在出发点右边)．如左边吸收壁距出发点 a 步，右边吸收壁距出发点 b 步，则当 S_n 第一次取值 a 时，表示质点已移动到左吸收壁，而当 S_n 第一次取值 $-b$ 时，表示质点游动到右吸收壁．

显然，我们不难把上述的随机游动及其模拟推广到以下情形：第一，边界点不一定是吸收壁，而是当质点到达边界点时只以一定的概率被吸收，同时还可以正的概率往回继续移动；第二，确定移动方向的概率 p 可以依赖于质点的位置．

下面再介绍一下二维（格子点上的)随机游动．质点从某点 O 出发，向左，向右，向上，向下移动一步的概率分别是 p_1，p_2,p_3,p_4，$p_1+p_2+p_3+p_4=1$．质点移动一步到达新的位置以后，又以同样的规律继续游动，直至碰到某边界点（在边界 Γ 上或距边界不足一步的点)后游动停止，参看图 11.2.2.

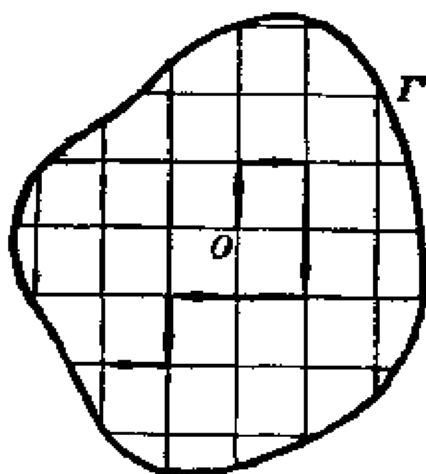


图 11.2.2

模拟方法是取(0,1)上均匀分布的随机数 r_1，若

$0<r_1<p_1$　表示质点左移一步；

$p_1\leqslant r_1<p_1+p_2$　表示质点右移一步；

$p_1+p_2\leqslant r_1<p_1+p_2+p_3$　表示质点上移一步；

$p_1+p_2+p_3\leqslant r_1<p_1+p_2+p_3+p_4$　表示质点下移一步.

依次取一列在(0,1)上均匀分布的随机数 $r_1,r_2,\cdots$，根据 $r_n(n=1,2,\cdots)$ 的值确定质点在第 n 步移动的位置，于是这一随机数列即可确定质点游动的轨迹，当质点到达某一边界点时游动停止.

最后，简单地讨论马尔科夫链(离散时间，可列状态)的模拟. 我们知道，随机游动是一种简单的马尔科夫链. 前者每一次状态转移只能到相邻的状态，而后者没有这个限制，它转移到每一可能状态的概率由转移概率矩阵给出. 因此，把上述随机游动的模拟方法作相应的推广就可以得到如下的马尔科夫链的模拟方法.

设马尔科夫链具有可列状态空间 $E=(0,1,2,\cdots)$，一步转移矩阵 $P=\{p_{ij}\}$，初始状态概率分布 $Q=(q_0,q_1,q_2,\cdots)$. 令

$$\sum_{j=0}^{n}p_{ij}=p_i^{(n)}\quad(i=0,1,2,\cdots),$$

$$\sum_{j=0}^{n}q_j=q^{(n)}.$$

取(0,1)上均匀分布的随机数 r_0，根据 r_0 确定初始状态，若 $q^{(n_0-1)}<r_0\leqslant q^{(n_0)}$，则令初始状态为 n_0. 再取(0,1)上均匀分布的随机数 r_1，据此确定第一次转移的状态，即若 $p_{n_0}^{(n_1-1)}<r_1\leqslant p_{n_0}^{(n_1)}$ 时说链从状态 n_0 转移到 n_1，如此继续下去就得到一状态序列 $n_0,n_1,n_2,\cdots$，它给出所考察的马尔科夫链的一个轨道(样本函数).

§11.3　定积分的概率计算方法

一、常用的两种算法

在 §11.1 节中我们已经介绍过一种最简单的积分例子，现在讨论一般的积分型式

$$I=\int_a^b f(x)\,dx, \tag{11.3.1}$$

积分域(a,b)有限，而且假设$f(x)$是(a, b)上的非负有界可积函数. 设

$$0\leqslant f(x)\leqslant c.$$

易见计算这个积分就是算出图 11.3.1 中有斜线的区域A的面积. 为此，可取n对随机数$(\xi_1,\eta_1)(\xi_2,\eta_2),\cdots,(\xi_n,\eta_n)$，其中$\xi_i(i=1,2,\cdots,n)$是区间$(a,b)$上均匀分布的随机数，$\eta_i(i=1,2,\cdots,n)$是区间$(0, c)$上均匀分布的随机数. 对每一对随机数$(\xi_i,\eta_i)$，考察$\xi_i$和$\eta_i$是否满足不等式

$$f(\xi_i)>\eta_i. \tag{11.3.2}$$

易见，满足不等式(11.3.2)的随机点就是落在区域A内的点，设共有m个，而全部随机点的点数目为n，则不难证明以概率 1 成立下面的近似公式:

$$\int_a^b f(x)\,dx\approx\frac{m}{n}c(b-a). \tag{11.3.3}$$

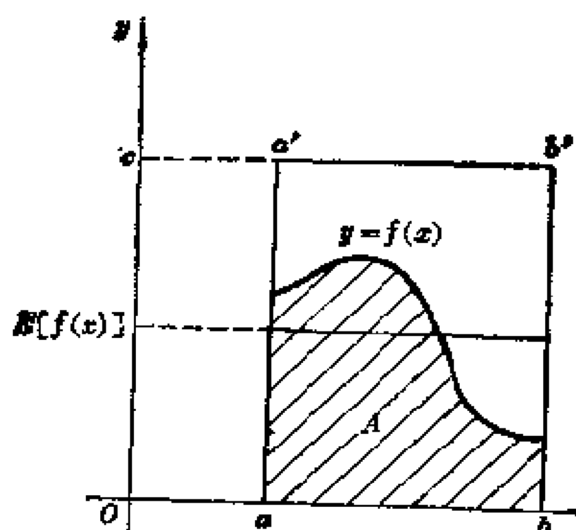


图 11.3.1

事实上，可以把上述算法设想为向矩形$a\leqslant x\leqslant b$，$0\leqslant y\leqslant c$中独立和均匀地掷点$P_i=(\xi_i,\eta_i)\ (i=1,2,\cdots,n)$. 设落于$A$中的点数为

$m(=m(n))$，则易见每次成功的概率$Q=\frac{A}{(b-a)c}$(即图 11.3.1 中有斜线的区域 A 和矩形 $a'abb'$ 面积之比)，又由强大数定律知以概率为 1 有

$$\lim_{n\to\infty}\frac{m}{n}=Q,$$

由此知当 n 足够大时近似式(11.3.3)以概率 1 成立。

因为 n 次掷点可看作是进行 n 次成功概率为 Q 的贝努里试验，m 是其中成功[即点落在 A 中，亦即不等式(11.3.2)成立]的次数，故 m 有二项分布 $b(n,\ Q)$，其平均值为 $E(m)=nQ$，方差为 $D(m)=nQ(1-Q)$。因此

$$\begin{aligned} E\left(\frac{m}{n}\right)&=\frac{1}{n}E(m)=Q,\\ D\left(\frac{m}{n}\right)&=\frac{1}{n^2}D(m)=\frac{Q(1-Q)}{n}. \end{aligned} \tag{11.3.4}$$

采用这种方法计算积分时，人们自然会问，对于任意给定的正数 ε 如果要求面积之计算值和真值之差

$$\left|\frac{m}{n}c(b-a)-\int_a^b f(x)dx\right|=\left|\frac{m}{n}c(b-a)-Qc(b-a)\right|$$

小于 ε 的概率不小于 $\alpha(0<\alpha<1)$，问 n 应取多大？根据中心极限定理，$\frac{m-nQ}{\sqrt{nQ(1-Q)}}$渐近地有正态分布 $N(0,1)$。因此问题化为求 n，使满足

$$P\left(\left|\frac{m}{n}c(b-a)-Qc(b-a)\right|<\varepsilon\right)=P\left(\left|\frac{m}{n}-Q\right|<\frac{\varepsilon}{c(b-a)}\right)$$

$$\approx\Phi\left(\frac{\varepsilon}{c(b-a)}\sqrt{\frac{n}{Q(1-Q)}}\right)-\Phi\left(-\frac{\varepsilon}{c(b-a)}\sqrt{\frac{n}{Q(1-Q)}}\right)\geqslant\alpha. \tag{11.3.5}$$

其中，把 ε、Q、α 作为已知数，即可求出满足(11.3.5)的最小 n 值。

实用中可以通过查正态分布表找出 $t_\alpha>0$ 使

$$\Phi(t_\alpha)-\Phi(-t_\alpha)\approx\alpha.$$

例如，当 $B=99.7\%$，95.6%和 68.3% 时，t_α 分别是 3，2 和 1. 找出 t_B 后，解(11.3.5)就变成解

$$\frac{\varepsilon}{c(b-a)}\sqrt{\frac{n}{Q(1-Q)}}=t_\alpha.$$

由此得

$$n\approx\frac{t_\alpha^2c^2(b-a)^2Q(1-Q)}{\varepsilon^2}. \tag{11.3.6}$$

但是，实际上我们并不知道 Q，因此必须首先设法估计 Q. 可用试算或其它办法得到 Q 的一个估计值 $\hat{Q}$，然后以 $\hat{Q}$ 值代入(11.3.6)式中进行计算.

上述计算方法的思想在前面 §11.1 中计算(0，1)上的积分时已用过，这就是利用随机变数在指定区域内取值的频率来计算积分，我们称之为频率法.

第二种计算积分的方法是利用随机变数的平均值来计算积分(11.3.1). 假设 ξ_i 是一列在区间(a,b)上均匀分布的独立随机变数，若函数$f(x)$在区间(a,b)上可积，则$\{f(\xi_i)\}$也是一列相互独立同分布的随机变数，算出它们的统计平均值并定义

$$F_n=\frac{1}{n}\sum_{i=1}^{n}f(\xi_i). \tag{11.3.7}$$

于是我们就有如下的近似计算公式

$$\int_a^b f(x)dx\approx(b-a)F_n\ (\text{记为 } I_n). \tag{11.3.8}$$

事实上，这时有

$$E[f(\xi_i)]=\frac{\int_a^b f(x)dx}{b-a}=\frac{I}{b-a}, \tag{11.3.9}$$

由强大数定律知以概率 1 成立

$$\lim_{n\to\infty}(b-a)\frac{\sum_{i=1}^{n}f(\xi_i)}{n}=(b-a)E[f(\xi_i)]=I.$$

因此，当 n 足够大时，可得近似公式(11.3.8).

我们还不难算出

$$E(I_n)=\frac{(b-a)}{n}\sum_{i=1}^{n}E[f(\xi_i)]=(b-a)E[f(\xi_i)]=I.$$

若 $f(x)$ 在 (a,b) 上平方可积，则方差

$$\begin{aligned}D(I_n)&=\frac{(b-a)^2}{n^2}\sum_{i=1}^{n}D[f(\xi_i)]\\&=\frac{(b-a)^2}{n}\{E[f^2(\xi_i)]-(E[f(\xi_i)])^2\}\\&=\frac{1}{n}\left[(b-a)\int_a^b f^2(x)dx-\left(\int_a^b f(x)dx\right)^2\right].\end{aligned}\tag{11.3.10}$$

由此并根据中心极限定理就可以按频率法那样讨论精确度和计算次数的问题，这时有近似公式

$$n\approx\frac{1}{\varepsilon^2}t_\alpha^2(b-a)^2\cdot D[f(\xi_i)],\tag{11.3.11}$$

其中符号 ε, t_α 的意义同前.

例11.3.1 用蒙特卡罗方法计算积分

$$I=\int_0^1\frac{e^x}{4}dx\tag{11.3.12}$$

先用频率法计算，为此从均匀分布的随机数表(本例所用随机数表见[35]P.484)中每次抽取一对数 δ_1 和 δ_2，例如，可以从该表中第 41个数字开始，挨次抽一个数作为 δ_1，从第 61 个数字开始，挨次抽一个数作为 δ_2，这时要用 10^{-5} 乘以表中的数字 [因附表是区间(0, 99999)内均匀分布的随机数，而我们的积分区间是(0,1)]，

并四舍五入取到小数点后第三位，然后算出 $f(\delta_1)=\frac{e^{\delta_1}}{4}$. 将它与 δ_2 比较，若 $f(\delta_1)>\delta_2$，认为试验成功. 设成功的次数为 m，则由(11.3.3)式知成功次数 m 除以总的试验次数 n，所得的商就是积分(11.3.12)式的近似值（这时 $b-a=1, c=1$）. 表11.3.1 列出了试验次数 n 等于 1 到20的计算结果 I'_n.

表 11.3.1

n	1	2	3	4	5	6	7	8	9	10
δ_1	0.667	0.993	0.242	0.940	0.610	0.131	0.352	0.646	0.646	0.680
δ_2	0.491	0.182	0.192	0.025	0.557	0.530	0.865	0.105	0.564	0.136
$f(\delta_1)$	0.487	0.675	0.318	0.640	0.460	0.285	0.356	0.477	0.477	0.493
m	0	1	2	3	3	3	3	4	4	5
I'_n	0	0.5	0.667	0.750	0.600	0.500	0.429	0.500	0.444	0.500
I''_n	0.487	0.581	0.493	0.530	0.516	0.478	0.460	0.416	0.423	0.430
n	11	12	13	14	15	16	17	18	19	20
δ_1	0.398	0.339	0.806	0.699	0.984	0.327	0.129	0.146	0.669	0.430
δ_2	0.579	0.541	0.238	0.432	0.033	0.462	0.284	0.161	0.990	0.547
$f(\delta_1)$	0.372	0.351	0.560	0.503	0.669	0.347	0.284	0.289	0.488	0.384
m	5	5	6	7	8	8	8	9	9	9
I'_n	0.455	0.417	0.462	0.500	0.533	0.500	0.471	0.500	0.470	0.450
I''_n	0.424	0.418	0.429	0.434	0.450	0.444	0.434	0.426	0.429	0.427

再用平均值方法来计算，这时只需要抽取随机数 δ_1 及计算 $f(\delta_1)$，然后按公式(11.3.8)算出积分近似值. 表11.3.1 中也列出了试验次数 n 等于 1 到 20 的计算结果 I''_n.

用分析方法算出积分(11.3.12)式的准确值是

$$I=\int_0^1\frac{e^x}{4}dx=\frac{1}{4}(e-1)=0.42957\cdots\approx0.430.$$

由此得到当 $n=20$ 时，频率法的计算误差为 4.7%. 平均值法的误差比较小，为 0.7%.

现在进一步问：如果要求以 99.7% 的概率，保证 I_n 作为 I 的近似值能准确到小数点后第三位，即保证

$$|I_n-I|<0.001,$$

那末 n 最低限度应取多大？

对于频率法，由公式(11.3.6)得

$$n\approx\frac{9\times0.43\times0.57}{(0.001)^2}=2.2059\times10^6.$$

对于平均值法，因为这时

$$D(I_n)=D\left(\frac{1}{4n}\sum_{i=1}^{n}e^{\xi_i}\right)=\frac{1}{n}\left[\frac{1}{16}\int_0^1 e^{2x}dx-\left(\frac{1}{4}\int_0^1 e^x dx\right)^2\right]$$

$$=\frac{1}{n}\left[\frac{e^2-1}{32}-\frac{(e-1)^2}{6}\right]=\frac{0.0151}{n}.$$

因为由 (11.3.10) 知 $D[f(\xi_i)]=\frac{n}{(b-a)^2}D(I_n)=0.0151$，故由(11.3.11)得 $n\approx\frac{9\times0.0151}{(0.001)^2}=0.1359\times10^6$.

将以上结果比较可看出，平均值法较频率法好，因为它的计算误差较小，或者说在同样精确度要求下，它所需的计算次数较小.

二、重积分的计算

上面介绍的两种算法原则上也适用于多重积分，这是蒙特卡罗方法的一个优点. 现就多重积分的情形简单地介绍一下平均值方法.

以 $x=(x_1,x_2,\cdots,x_m)$ 表示 m 维空间 R^m 中的点，$f(x)$ 是 m 元函数，它在有界区域 D 内可积，记

$$I=\int_D f(x)dx=\int\cdots\int_D f(x_1,x_2,\cdots,x_m)dx_1\cdots dx_m, \qquad (11.3.13)$$

现计算 I.

取一维区间 (a,b)，使得

$$D\subset \underbrace{(a,b)\times(a,b)\times\cdots\times(a,b)}_{m个}=(a,b)^m.$$

取 m 个相互独立且在 (a,b) 中均匀分布的随机变数 $\xi_1,\xi_2,\cdots,\xi_m$，令 $\xi=(\xi_1,\xi_2,\cdots,\xi_m)$. 可以证明，在 $\xi\in D$ 的条件下，ξ 在 D 中均匀分布，这就是说有

$$P\{\xi\in V\mid\xi\in D\}=\frac{|V|}{|D|}. \qquad (11.3.14)$$

其中 V 是 D 的任一子域，$|V|$ 和 $|D|$ 分别表示 V 和 D 的 m 维体积. 参看图 11.3.2($m=2$ 的情形).

事实上，这时有

$$P(\xi\in V\mid\xi\in D)=\frac{P(\xi\in V)}{P(\xi\in D)}$$

$$=\frac{|V|}{(b-a)^m}\bigg/\frac{|D|}{(b-a)^m}$$

$$=\frac{|V|}{|D|}.$$

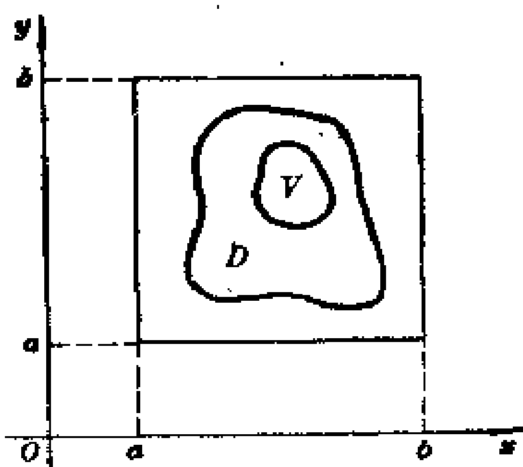


图 11.3.2

由此进一步推得

$$E[f(\xi)\mid\xi\in D]=\frac{\int_D f(x)\,dx}{|D|}=\frac{I}{|D|}. \qquad (11.3.15)$$

根据上式，可得到以下计算方法：如同上面的 ξ 那样，构造一列相互独立的 m 维随机向量 $\{\xi^{(i)}\}$，$\xi^{(i)}=(\xi_1^{(i)},\xi_2^{(i)},\cdots,\xi_m^{(i)})\xi^{(i)}$ 在 m 维立方体 $(a,b)^m$ 中均匀分布. 如果 $\xi^{(i)}\in D$，就选出 $\xi^{(i)}$，如果 $\xi^{(i)}\overline{\in}D$ 则舍弃之. 如此共选出 n 个 $\xi^{(i)}$，记为 $\xi^{(i_1)},\xi^{(i_2)},\cdots,\xi^{(i_n)}$，由 (11.3.15) 式得到以概率 1 成立

$$|D|\lim_{n\to\infty}\frac{\sum_{k=1}^{n}f(\xi^{(i_k)})}{n}=I,$$

于是我们就可以取

$$I_n=|D|\ \frac{1}{n}\sum_{k=1}^{n}f(\xi^{(i_k)}) \qquad (11.3.16)$$

作为积分 I 的近似值.

最后要指出一点，在实际计算中，最好把分析计算方法和概率计算方法结合起来. 这就是说，在(11.3.13)式中尽可能选用数学分析方法直接算出对某些变元的积分，以降低整个积分的重数，对于剩余的无法直接计算的变元再用平均值法，这样做往往能减少计算量，提高计算效率.

§11.4 某些方程的概率解法

一、线性方程组的求解

我们讨论线性方程组

$$\sum_{k=1}^{n}a_{ik}x_k=b_i\ \ (i=1,2,\cdots,n) \qquad (11.4.1)$$

的求解. 首先考虑二次型

$$Q=Q(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{n}\alpha_i\Big(\sum_k a_{ik}x_k-b_i\Big)^2, \quad (11.4.2)$$

这里 $\alpha_i>0$ 是常数. 易见求(11.4.1)式的解等价于求 Q 的最小点 $(x_1^0,x_2^0,\cdots,x_n^0)$. 对于任意常数 $A>0$，易见

$$E:\ Q(x_1,x_2,\cdots,x_n)\leqslant A$$

是一个 n 维椭球，这个椭球的中心由 $\sum_k a_{ik}x_k-b_i=0$ $(i=1,2,\cdots,n)$ 确定，记为 $(x_1^0,x_2^0,\cdots,x_n^0)$. 每个通过中心的 n 维超平面

$x_s=x_s^0$ 都把椭球分成(n维)体积相等的两部分，我们的解法就是基于这一几何性质，即求$x_1^0, x_2^0, \cdots, x_n^0$，使椭球$E$属于$x_s \leqslant x_s^0$那部分恰有一半的体积。

作n维立方体$(a,b)^n$，使得椭球$E \subset (a,b)^n$。设$\xi_1, \cdots, \xi_n$是n个在(a,b)上均匀分布的独立随机变数，则$\xi=(\xi_1, \cdots, \xi_n)$是$n$维随机变数，由上节关于多重积分的讨论知道，在$\xi \in E$的条件下，$\xi$在$E$中有均匀分布。取$\xi$的$l$个独立的子样值$\xi^{(1)}, \cdots, \xi^{(l)}$，如果$\xi^{(i)} \in E$就选取$\xi^{(i)}$，否则舍弃之。设共选出$m$个这样的样本值$\xi^{(i_1)}, \xi^{(i_2)}, \cdots, \xi^{(i_m)}$ $(m \leqslant l)$，每一$\xi^{(i_k)}$ $(k=1, 2, \cdots, m)$的分量表示为

$$\xi^{(i_k)}=(\xi_1^{(i_k)}, \cdots, \xi_n^{(i_k)}).$$

因为$\xi^{(i_k)}$在E中均匀分布，而超平面$x_s=x_s^0$把椭球分成体积相等的两部分，故当m充当大时，$\xi^{(i_1)}, \xi^{(i_2)}, \cdots, \xi^{(i_m)}$中应差不多各有一半分别在此超平面的两侧，即这些向量的第s个分量$\xi_s^{(i_1)}, \xi_s^{(i_2)}, \cdots, \xi_s^{(i_m)}$中应约有一半不大于(或不小于)$x_s^0$。把这些分量按大小排列如下：

$$\xi_s^{(1)} \leqslant \xi_s^{(2)} \leqslant \cdots \leqslant \xi_s^{(m)}, \qquad (11.4.3)$$

其中的$\xi_s^{\left(\left[\frac{m}{2}\right]+1\right)}$ $\left(\left[\frac{m}{2}\right]\text{表示}\frac{m}{2}\text{的整数部分}\right)$自然应最靠近$x_s^0$，因此我们可以取

$$\xi_s^{\left(\left[\frac{m}{2}\right]+1\right)} \approx x_s^0 \qquad (s=1, 2, \cdots, n), \qquad (11.4.4)$$

即以$\left(\xi_1^{\left(\left[\frac{m}{2}\right]+1\right)}, \xi_2^{\left(\left[\frac{m}{2}\right]+1\right)}, \cdots, \xi_n^{\left(\left[\frac{m}{2}\right]+1\right)}\right)$作为方程组(11.4.1)的近似解。

以上是取子样(11.4.3)式的中数作为x_s^0的近似值，我们同样也可以取它的平均值作近似，即

$$\frac{1}{m}\sum_{k=1}^{m}\xi_s^{(k)} = \frac{1}{m}\sum_{k=1}^{m}\xi_s^{(i_k)} \approx x_s^0. \qquad (11.4.5)$$

现在对一类满足某种条件的线性方程组给出另一种解法.

先把(11.4.1)式改写成

$$\sum_{s=1}^{n}(\delta_{is}-h_{is})x_s=b_i \quad (i=1,2,\cdots,n), \qquad (11.4.6)$$

其中

$$\delta_{is}=\begin{cases}1, & i=s;\\ 0, & i\neq s.\end{cases}$$

易见(11.4.1)式和(11.4.6)式的系数有如下关系

$$a_{is}=\begin{cases}1-h_{is}, & i=s;\\ -h_{is}, & i\neq s.\end{cases}$$

如果把(10.4.6)式写成矩阵方程:

$$(I-H)X=B, \qquad (11.4.7)$$

其中 I 为单位矩阵, H,B 是已知的系数矩阵, X 为待求的未知向量. 令

$$\|H\|=\max_i\left(\sum_{s=1}^{n}|h_{is}|\right),$$

我们在 $\|H\|<1$ 的条件之下给出方程(11.4.7)式的一种解法. 由条件 $\|H\|<1$ 得

$$X=(I-H)^{-1}B=(I+H+H^2+\cdots)B. \qquad (11.4.8)$$

于是 X 的第 i 个分量 x_i 为

$$x_i=b_i+(HB)_i+(H^2B)_i+\cdots$$

$$=b_i+\sum_{i_1=1}^{n}h_{ii_1}b_{i_1}+\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}h_{ii_1}h_{i_1i_2}b_{i_2}+\cdots$$

$$=b_i+\sum_{k=1}^{\infty}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n}h_{ii_1}h_{i_1i_2}\cdots h_{i_{k-1}i_k}b_{i_k}. \qquad (11.4.9)$$

现在要定义一个马尔科夫链，使它的某一数学特征恰好等于 x_i. 任取一矩阵 $P=(p_{is})$,) $(i,s=1,2,\cdots,n)$, 它满足条件

$$p_{is}\geqslant 0;\qquad \sum_s p_{is}\leqslant 1; \tag{11.4.10}$$

$$p_{is}>0,\qquad 如果\ h_{is}\neq 0. \tag{11.4.11}$$

令

$$p_i=1-\sum_s p_{is}\qquad (如\ b_i\neq 0, 则还应取\ p_i>0). \tag{11.4.12}$$

考虑一具有以下性质的马尔科夫链：如果它从状态 i 出发，则它下一步转移到状态 s 的概率为 p_{is}，而永远停止运动的概率是 p_i. 现设它从状态 i 出发，经 k 步后停止运动，则其轨道 J 可记为

$$J:\ i\to i_1\to i_2\to\cdots\to i_k\ (k\geqslant 0, i_0=i). \tag{11.4.13}$$

由定义知它有轨道(11.4.13)的概率是

$$p_{ii_1}p_{i_1i_2}\cdots p_{i_{k-1}i_k}p_{i_k}. \tag{11.4.14}$$

又在轨道 J 上定义函数

$$V(J)=\begin{cases}\dfrac{h_{ii_1}}{p_{ii_1}}\cdot\dfrac{h_{i_1i_2}}{p_{i_1i_2}}\cdots\dfrac{h_{i_{k-1}i_k}}{p_{i_{k-1}i_k}}\cdot\dfrac{b_{i_k}}{p_{i_k}} & 如果\quad k>0;\\ b_i/p_i & \qquad\quad k=0.\end{cases} \tag{11.4.15}$$

以 E_i 表示从 i 出发时的条件数学期望，则由(11.4.14)，(11.4.15)两式得

$$\begin{aligned}E_i[V(J)]&=\frac{b_i}{p_i}\cdot p_i+\sum_{i_1}\frac{h_{ii_1}b_{i_1}}{p_{ii_1}p_{i_1}}p_{ii_1}p_{i_1}\\&+\sum_{i_1}\sum_{i_2}\frac{h_{ii_1}h_{i_1i_2}b_{i_2}}{p_{ii_1}p_{i_1i_2}p_{i_2}}p_{ii_1}p_{i_1i_2}p_{i_2}+\cdots\\&=b_i+\sum_{k=1}^{\infty}\sum_{i_1}\sum_{i_2}\cdots\sum_{i_k}h_{ii_1}h_{i_1i_2}\cdots h_{i_{k-1}i_k}b_{i_k}.\end{aligned}$$

将此式和(11.4.9)式比较即得

$$x_i=E_i[V(J)]. \tag{11.4.16}$$

于是，我们得到如下的计算步骤：

1° 用§11.2所介绍的方法，模拟转移概率矩阵$P=(p_{ik})$和初始状态为 i 的马尔科夫链；

2° 独立地造出 n 个形如(11.4.13)的轨道 J_1，$J_2,\cdots$，J_n 并按(11.4.15)算出 $V(J_1),V(J_2),\cdots,V(J_n)$；

3° 以$\frac{1}{n}\sum_{k=1}^{n}V(J_k)$作为 x_i 的估计值.

本法优点是每个 x_i 可单独求出而不需要同时算出其它的 x_j. 同时，这方法的原理还可应用于求解某些积分方程.

二、一些偏微分方程的求解

我们首先讨论一个最典型的问题：二维拉普拉斯方程 $\Delta u=0$ $\left(\text{即}\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0\right)$的狄利赫里问题. 即设有某个区域 D，在其边界 Γ 上给定某个函数 $f(Q)$，试求函数 $u(P)$，使得它在给定区域 D 的内部满足拉普拉斯方程

$$\Delta u=0, \tag{11.4.17}$$

在边界 Γ 上取给定函数值

$$u|_\Gamma=f(Q), \tag{11.4.18}$$

其中 $Q\in\Gamma$.

这个问题可以化为如下的有限差分问题：给定步长 h，并在所考虑的平面上作正方形网格，我们只研究落在区域 D 上的网格点，网格点分成两类：一类是它的四个相邻网格点位于区域内部或边界 Γ 上，这类点称做内部网格点，在图 11.4.1 中以△标出. 另一类是它的四个相邻网格点不全在区域内部或边界上，这类点称做边界网格点，在图 11.4.1 中以○标出. 在内部网格点 P 上可以建立相应于原来的微分方程组的有限差分方程组

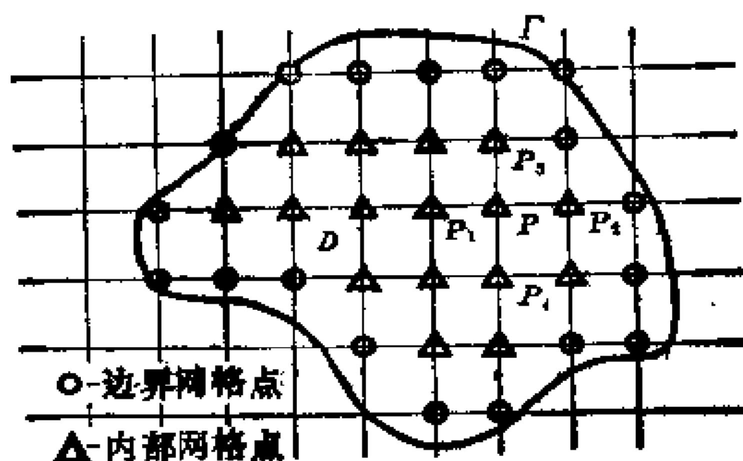


图 11.4.1

$$u(P)=\frac{1}{4}[u(P_1)+u(P_2)+u(P_3)+u(P_4)], \quad (11.4.19)$$

其中 P_1, P_2, P_3, P_4 是内点 P 的四个相邻网格点. 对于边界网格点 Q 上的取值，则按照某种规则利用给定在区域边界 Γ 上的函数 f 来确定，记为

$$u(Q)=f(Q). \quad (11.4.20)$$

现在来研究模拟方程组(11.4.19)及边界条件(11.4.20)的概率模型. 考虑一个质点在所有内部网格点和边界网格点上作随机游动①: 质点可从区域内任一网格点 P 出发，以相等的概率(即各等于1/4)游动到任一个相邻的网格点，而当它游动到某个内部网格点后，又继续以相同的概率游动到它的四个相邻网格点中的一点，如此下去，一直到质点游动到某一边界点时就停止在那里不再游动，这就产生如何求质点从内网格点 P 出发，最终游动到边界网格点 Q 的概率 $u(P,Q)$ 的问题. 要推出这个概率的明显表示式，一般是复杂和困难的. 但是，我们的问题只需推出有关这概率的某些关系式即可. 同时概率论中已经证明，质点经过有限步游动后最终停在某个边界点的概率等于 1 (参看[13])，这保证质点实际上不会无

① 这一模型在国外有关文献中常称做“醉汉问题”.

休止地在区域内部游动．易见事件“质点从P游动到Q”是以下述四个不相容事件之和：“质点从P游动到P_1，然后从P_1游动到Q”；“质点从P游动到P_2，然后从P_2游动到Q”；“质点从P游动到P_3，然后从P_3游动到Q”；“质点从P游动到P_4，然后从P_4游动到Q”．因此，按全概率公式有

$$u(P,Q)=\frac{1}{4}[u(P_1,Q)+u(P_2,Q)+u(P_3,Q)+u(P_4,Q)]. \tag{11.4.21}$$

易见(11.4.21)式和(11.4.19)式是一致的，此外，由模型本身的性质知概率$u(P,Q)$满足下面的边界条件

$$u(Q',Q)=\begin{cases}1, & \text{当 } Q'=Q;\\ 0, & \text{当 } Q'\neq Q,\end{cases} \tag{11.4.22}$$

这里Q'和Q皆为边界网格点．

已经知道，在给定的边界条件下存在满足方程(11.4.19)式的唯一解．

假定模拟质点游动N次，每次皆从P出发，如果其中质点游动最终到达边界点Q的试验次数为M，则频率$\frac{M}{N}$给出满足边界条件(11.4.22)的方程(11.4.21)的近似解，即$u(P,Q)\approx\frac{M}{N}$.

如果我们还规定当质点游动到边界网格点Q后，我们要付出代价$f(Q)$，则由P发出的质点游动所要付出的代价是一随机变数，我们用$\xi(P)$表之．设共有s个边界网格点$Q_i(i=1,2,\cdots,s)$．于是我们为质点游动付出代价$f(Q_i)$的概率为$u(P,Q_i)$，因而付出代价的平均值为

$$W(P)=E[\xi(P)]=\sum_{i=1}^{s}f(Q_i)u(P,Q_i) \tag{11.4.23}$$

显然$W(P)$依赖于出发点P，容易验证这函数也满足差分方程

$$W(P)=\frac{1}{4}[W(P_1)+W(P_2)+W(P_3)-W(P_4)], \quad (11.4.24)$$

同时还满足所要求的边界条件.事实上，在(11.4.23)式中令$P=Q$,则由(11.4.22)式可得$W(Q)=f(Q)$，于是函数$W(P)$就是我们所要求的解.

以上是就二维的情形加以讨论，这方法还可以应用于多维情形和更一般的边界问题. 我们可以证明，如果所考虑的区域为边长等于r的n维立方体，即变量$x_1,x_2,\cdots,x_n$在n维立方体:$0\leqslant x_i\leqslant r(i=1,2,\cdots,n)$内取值，则质点从区域内网格点$P$出发游动直至到达边界网格点$Q$为止所走的步数的平均值$z(P)$满足不等式

$$z(P)<r^2/2.$$

这就是说,质点游动的时间不依赖于网格的维数,这是蒙特卡罗方法的一个重要特点,即它较之一般方法更适用于高维的情形.

对于更一般的二阶线性椭圆型方程

$$a(x,y)\frac{\partial^2 u}{\partial x^2}+b(x,y)\frac{\partial^2 u}{\partial x\partial y}+c(x,y)\frac{\partial^2 u}{\partial y^2}+d(x,y)\frac{\partial u}{\partial x}$$

$$+e(x,y)\frac{\partial u}{\partial y}+f(x,y)u=0, \quad (11.4.25)$$

(其中$ac-b^2>0$),它的解可以利用推广了的随机游动模型，即质点从某一个网格点游动到另一个网格点的转移概率可以不相等，它依赖于质点当时所在的网格点.

我们还可以研究更一般的边界条件，例如

$$\left.\frac{\partial u}{\partial n}\right|_\Gamma=ku|_\Gamma+f(Q),$$

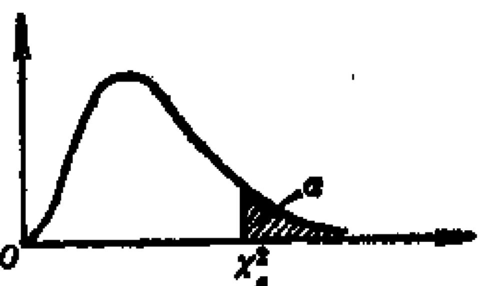
这时用来求解的游动模型应当假定质点到达边界点后能以正概率返回区域内继续游动,而不一定停止在该边界点上.

至于其它类型的偏微分方程，其中有一些也能用类似的方法求解.

附表

表 1. χ^2-分布的上侧临界值表

$$P(\chi_f^2 > \chi_\alpha^2) = \alpha$$

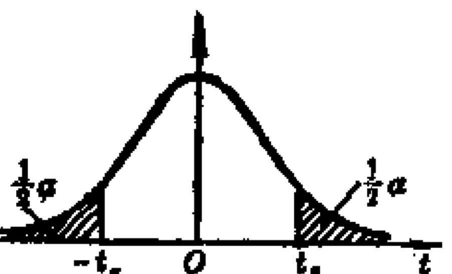


f \ α	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001	α / f
1	0.0^3157	0.0^3628	0.0^2393	0.0158	0.0642	0.148	0.455	1.074	1.642	2.706	3.841	5.412	6.635	10.828	1
2	0.0201	0.0404	0.103	0.211	0.446	0.713	1.386	2.408	3.219	4.605	5.991	7.824	9.210	13.816	2
3	0.115	0.185	0.352	0.584	1.005	1.424	2.366	3.665	4.624	6.251	7.815	9.837	11.345	16.266	3
4	0.297	0.429	0.711	1.064	1.649	2.195	3.357	4.878	5.989	7.779	9.488	11.668	12.277	18.467	4
5	0.554	0.752	1.145	1.610	2.343	3.000	4.351	6.064	7.289	9.236	11.070	13.388	15.068	20.515	5
6	0.872	1.134	1.635	2.204	3.070	3.828	5.348	7.231	8.558	10.645	12.592	15.033	16.812	22.458	6
7	1.239	1.564	2.167	2.833	3.822	4.671	6.346	8.383	9.803	12.017	14.067	16.622	18.475	24.322	7
8	1.646	2.032	2.733	3.490	4.594	5.527	7.344	9.524	11.030	13.362	15.507	18.168	20.090	26.125	8
9	2.088	2.532	3.325	4.168	5.380	6.393	8.343	10.656	12.242	14.684	16.919	19.679	21.666	27.877	9
10	2.558	3.059	3.940	4.865	6.179	7.267	9.342	11.781	13.442	15.987	18.307	21.161	23.209	29.588	10

11	3.053	3.609	4.575	5.578	6.989	8.148	10.341	12.899	14.631	17.275	19.675	22.618	24.725	31.264	11
12	3.571	4.178	5.226	6.304	7.807	9.034	11.340	14.011	15.812	18.549	21.026	24.054	26.217	32.909	12
13	4.107	4.765	5.892	7.042	8.634	9.926	12.340	15.119	16.985	19.812	22.362	25.472	27.688	34.528	13
14	4.660	5.368	6.571	7.790	9.467	10.821	13.339	16.222	18.151	21.064	23.685	26.873	29.141	36.123	14
15	5.229	5.985	7.261	8.547	10.307	11.721	14.339	17.322	19.311	22.307	24.996	28.259	30.578	37.697	15
16	5.812	6.614	7.962	9.312	11.152	12.624	15.338	18.418	20.465	23.542	26.296	29.633	32.000	39.252	16
17	6.408	7.255	8.672	10.085	12.002	13.531	16.338	19.511	21.615	24.769	27.587	30.995	33.409	40.790	17
18	7.015	7.906	9.390	10.865	12.857	14.440	17.338	20.601	22.760	25.989	28.869	32.346	34.805	42.312	18
19	7.633	8.567	10.117	11.651	13.716	15.352	18.338	21.689	23.900	27.204	30.144	33.687	36.191	43.820	19
20	8.260	9.237	10.851	12.443	14.578	16.266	19.337	22.775	25.038	28.412	31.410	35.020	37.566	45.315	20
21	8.897	9.915	11.591	13.240	15.445	17.182	20.337	23.858	26.171	29.615	32.671	36.343	38.932	46.797	21
22	9.542	10.600	12.338	14.041	16.314	18.101	21.337	24.939	27.301	30.813	33.924	37.659	40.289	48.268	22
23	10.196	11.293	13.091	14.848	17.187	19.021	22.337	26.018	28.429	32.007	35.172	38.968	41.638	49.728	23
24	10.856	11.992	13.848	15.659	18.062	19.943	23.337	27.096	29.553	33.196	36.415	40.270	42.980	51.179	24
25	11.524	12.697	14.611	16.473	18.940	20.867	24.337	28.172	30.675	34.382	37.652	41.566	44.314	52.618	25
26	12.198	13.409	15.379	17.292	19.820	21.792	25.336	29.246	31.795	35.563	38.885	42.856	45.642	54.052	26
27	12.879	14.125	16.151	18.114	20.703	22.719	26.336	30.319	32.912	36.741	40.113	44.140	46.963	55.476	27
28	13.565	14.847	16.928	18.939	21.588	23.647	27.336	31.391	34.027	37.916	41.337	45.419	48.278	56.893	28
29	14.256	15.574	17.708	19.768	22.475	24.577	28.336	32.461	35.139	39.087	42.557	46.693	49.588	58.301	29
30	14.953	16.306	18.493	20.599	23.364	25.508	29.336	33.530	36.250	40.256	43.773	47.962	50.892	59.703	30

表 2. t-分布的双侧临界值表

$$P(|t|>t_{\alpha})=\alpha$$



f \ α	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001	α / f
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619	1
2	.142	.289	.445	.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598	2
3	.137	.277	.424	.584	.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924	3
4	.134	.271	.414	.569	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610	4
5	.132	.267	.408	.559	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859	5
6	.131	.265	.404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959	6
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405	7
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041	8
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781	9
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587	10
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437	11
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318	12
13	.128	.259	.394	.538	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221	13
14	.128	.258	.393	.537	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140	14
15	.128	.258	.393	.536	.691	.866	1.074	1.341	1.753	2.131	3.602	2.947	4.073	15

16	.128	.258	.392	.535	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015	16
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965	17
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922	18
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883	19
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850	20
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819	21
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792	22
23	.127	256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767	23
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745	24
25	.127	.256	.390	.531	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725	25
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707	26
27	.127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690	27
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674	28
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659	29
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646	30
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551	40
60	.126	.254	.387	.527	.679	.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460	60
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373	120
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291	∞

表 3. F 检验的临界值(F_α)表

$$P(F>F_\alpha)=\alpha$$

$\alpha=0.10$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞	n_1 / n_2
1	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	61.2	61.7	62.3	62.7	63.0	63.2	63.3	63.3	1
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.42	9.44	9.46	9.47	9.48	9.49	9.49	9.49	2
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.20	5.18	5.17	5.15	5.14	5.14	5.14	5.13	3
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.87	3.84	3.82	3.80	3.78	3.77	3.76	3.76	4
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.24	3.21	3.17	3.15	3.13	3.12	3.11	3.10	5
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.87	2.84	2.80	2.77	2.75	2.73	2.73	2.72	6
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.63	2.59	2.56	2.52	2.50	2.48	2.48	2.47	7
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.46	2.42	2.38	2.35	2.32	2.31	2.30	2.29	8
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.34	2.30	2.25	2.22	2.19	2.17	2.17	2.16	9
10	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.24	2.20	2.16	2.12	2.09	2.07	2.06	2.06	10
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.17	2.12	2.08	2.04	2.00	1.99	1.98	1.97	11
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.10	2.06	2.01	1.97	1.94	1.92	1.91	1.90	12
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.05	2.01	1.96	1.92	1.88	1.86	1.85	1.85	13
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.01	1.96	1.91	1.87	1.83	1.82	1.80	1.80	14
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.97	1.92	1.87	1.83	1.79	1.77	1.76	1.76	15

16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.08	1.94	1.89	1.84	1.79	1.76	1.74	1.73	1.72	16
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.91	1.86	1.81	1.76	1.73	1.71	1.69	1.69	17
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.89	1.84	1.78	1.74	1.70	1.68	1.67	1.66	18
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.86	1.81	1.76	1.71	1.67	1.65	1.64	1.63	19
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.84	1.79	1.74	1.69	1.65	1.63	1.62	1.61	20
22	2.95	2.56	2.36	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.81	1.76	1.70	1.65	1.61	1.59	1.58	1.57	22
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.78	1.73	1.67	1.62	1.58	1.56	1.54	1.53	24
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.76	1.71	1.65	1.59	1.55	1.53	1.51	1.50	26
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.74	1.69	1.63	1.57	1.53	1.50	1.49	1.48	28
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.72	1.67	1.61	1.55	1.51	1.48	1.47	1.46	30
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.66	1.61	1.54	1.48	1.43	1.41	1.39	1.38	40
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.63	1.57	1.50	1.44	1.39	1.36	1.34	1.33	50
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.60	1.54	1.48	1.41	1.36	1.33	1.31	1.29	60
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68	1.57	1.51	1.44	1.38	1.32	1.28	1.26	1.24	80
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.70	1.66	1.56	1.49	1.42	1.35	1.29	1.26	1.23	1.21	100
200	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.52	1.46	1.38	1.31	1.24	1.20	1.17	1.14	200
500	2.72	2.31	2.10	1.96	1.86	1.79	1.73	1.68	1.64	1.61	1.50	1.44	1.36	1.28	1.21	1.16	1.12	1.09	500
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.49	1.42	1.34	1.26	1.18	1.13	1.08	1.00	∞

$\alpha=0.05$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	n_1 / n_2
1	161	200	216	225	230	234	237	239	241	242	244	245	246	247	248	1
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	2
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66	3
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80	4
5	6.61	5.79	5.41	6.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56	5
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87	6
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44	7
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15	8
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94	9
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77	10
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65	11
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54	12
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46	13
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39	14
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33	15
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28	16
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23	17
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19	18
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16	19
20	4.35	3.49	3.10	2.87	2.21	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12	20
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10	21
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07	22
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.07	2.05	23
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03	24
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01	25

26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99	26
27	1.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97	27
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96	28
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94	29
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93	30
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14	2.07	2.01	1.97	1.94	1.91	32
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17	2.12	2.05	1.99	1.95	1.92	1.89	34
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15	2.11	2.03	1.98	1.93	1.90	1.87	36
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.02	1.96	1.92	1.88	1.85	38
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84	40
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	1.99	1.93	1.89	1.86	1.83	42
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	1.98	1.92	1.88	1.84	1.81	44
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09	2.04	1.97	1.91	1.87	1.83	1.80	46
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08	2.03	1.96	1.90	1.86	1.82	1.79	48
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78	50
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75	60
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70	80
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68	100
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.96	1.91	1.83	1.77	1.72	1.69	1.65	125
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64	150
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62	200
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61	300
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59	500
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58	1000
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.69	1.64	1.60	1.57	∞

$\alpha = 0.05$

n_2 \ n_1	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞	n_1 / n_2
1	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254	1
2	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	2
3	8.65	8.64	8.63	8.62	8.62	8.60	8.59	8.59	8.58	8.57	8.56	8.55	8.54	8.53	8.53	3
4	5.79	5.77	5.76	5.75	5.75	5.73	5.72	5.71	5.70	5.69	5.67	5.66	5.65	5.64	5.63	4
5	4.54	4.53	4.52	4.50	4.50	4.48	4.46	4.45	4.44	4.43	4.41	4.41	4.39	4.37	4.37	5
6	3.86	3.84	3.83	3.82	3.81	3.79	3.77	3.76	3.75	3.74	3.72	3.71	3.69	3.68	3.67	6
7	3.43	3.41	3.40	3.39	3.38	3.36	3.34	3.33	3.32	3.30	3.29	3.27	3.25	3.24	3.23	7
8	3.13	3.12	3.10	3.09	3.08	3.06	3.04	3.03	3.02	3.01	2.99	2.97	2.95	2.94	2.93	8
9	2.92	2.90	2.89	2.87	2.86	2.84	2.83	2.81	2.80	2.79	2.77	2.76	2.73	2.72	2.71	9
10	2.75	2.74	2.72	2.71	2.70	2.68	2.66	2.65	2.64	2.62	2.60	2.59	2.56	2.55	2.54	10
11	2.63	2.61	2.59	2.58	2.57	2.55	2.53	2.52	2.51	2.49	2.47	2.46	2.43	2.42	2.40	11
12	2.52	2.51	2.49	2.48	2.47	2.44	2.43	2.41	2.40	2.38	2.36	2.35	2.32	2.31	2.30	12
13	2.44	2.42	2.41	2.39	2.38	2.36	2.34	2.33	2.31	2.30	2.27	2.26	2.23	2.22	2.21	13
14	2.37	2.35	2.33	2.32	2.31	2.28	2.27	2.25	2.24	2.22	2.20	2.19	2.16	2.14	2.13	14
15	2.31	2.29	2.27	2.26	2.25	2.22	2.20	2.19	2.18	2.16	2.14	2.12	2.10	2.08	2.07	15
16	2.25	2.24	2.22	2.21	2.19	2.17	2.15	2.14	2.12	2.11	2.08	2.07	2.04	2.02	2.01	16
17	2.21	2.19	2.17	2.16	2.15	2.12	2.10	2.09	2.08	2.06	2.03	2.02	1.99	1.97	1.96	17
18	2.17	2.15	2.13	2.12	2.11	2.08	2.06	2.05	2.04	2.02	1.99	1.98	1.95	1.93	1.92	18
19	2.13	2.11	2.10	2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.91	1.89	1.88	19
20	2.10	2.08	2.07	2.05	2.04	2.01	1.99	1.98	1.97	1.95	1.92	1.91	1.88	1.86	1.84	20
21	2.07	2.05	2.04	2.02	2.01	1.98	1.96	1.95	1.94	1.92	1.89	1.88	1.84	1.82	1.81	21
22	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.92	1.91	1.89	1.86	1.85	1.82	1.80	1.78	22
23	2.02	2.00	1.99	1.97	1.96	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.79	1.77	1.76	23
24	2.00	1.98	1.97	1.95	1.94	1.91	1.89	1.88	1.86	1.84	1.82	1.80	1.77	1.75	1.73	24
25	1.98	1.96	1.95	1.93	1.92	1.89	1.87	1.86	1.84	1.82	1.80	1.78	1.75	1.73	1.71	25

26	1.97	1.95	1.93	1.91	1.90	1.87	1.85	1.84	1.82	1.80	1.78	1.76	1.73	1.71	1.69	26
27	1.95	1.93	1.91	1.90	1.88	1.86	1.84	1.82	1.81	1.79	1.76	1.74	1.71	1.69	1.67	27
28	1.93	1.91	1.90	1.88	1.87	1.84	1.82	1.80	1.79	1.77	1.74	1.73	1.69	1.67	1.65	28
29	1.92	1.90	1.88	1.87	1.85	1.83	1.81	1.79	1.77	1.75	1.73	1.71	1.67	1.65	1.64	29
30	1.91	1.89	1.87	1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.71	1.70	1.66	1.64	1.62	30
32	1.88	1.86	1.85	1.83	1.82	1.79	1.77	1.75	1.74	1.71	1.69	1.67	1.63	1.61	1.59	32
34	1.86	1.84	1.82	1.80	1.80	1.77	1.75	1.73	1.71	1.69	1.66	1.65	1.61	1.59	1.57	34
36	1.85	1.82	1.81	1.79	1.78	1.75	1.73	1.71	1.69	1.67	1.64	1.62	1.59	1.56	1.56	36
38	1.83	1.81	1.79	1.77	1.76	1.73	1.71	1.69	1.68	1.65	1.62	1.61	1.57	1.54	1.53	38
40	1.81	1.79	1.77	1.76	1.74	1.72	1.69	1.67	1.66	1.64	1.61	1.59	1.55	1.53	1.51	40
42	1.80	1.78	1.76	1.74	1.73	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.53	1.51	1.49	42
44	1.79	1.77	1.75	1.73	1.72	1.69	1.67	1.65	1.63	1.61	1.58	1.56	1.52	1.49	1.48	44
46	1.78	1.76	1.74	1.72	1.71	1.68	1.65	1.64	1.62	1.60	1.57	1.55	1.51	1.48	1.46	46
48	1.77	1.75	1.73	1.71	1.70	1.67	1.64	1.62	1.61	1.59	1.56	1.54	1.49	1.47	1.45	48
50	1.76	1.74	1.72	1.70	1.69	1.66	1.63	1.61	1.60	1.58	1.54	1.52	1.48	1.46	1.44	50
60	1.72	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.56	1.53	1.50	1.48	1.44	1.41	1.39	60
80	1.68	1.65	1.63	1.62	1.60	1.57	1.54	1.52	1.51	1.48	1.45	1.43	1.38	1.35	1.32	80
100	1.65	1.63	1.61	1.59	1.57	1.54	1.52	1.49	1.48	1.45	1.41	1.39	1.34	1.31	1.28	100
125	1.63	1.60	1.58	1.57	1.55	1.52	1.49	1.47	1.45	1.42	1.39	1.36	1.31	1.27	1.25	125
150	1.61	1.59	1.57	1.55	1.53	1.50	1.48	1.45	1.44	1.41	1.37	1.34	1.29	1.25	1.22	150
200	1.60	1.57	1.55	1.53	1.52	1.48	1.46	1.43	1.41	1.39	1.35	1.32	1.26	1.22	1.19	200
300	1.58	1.55	1.53	1.51	1.50	1.46	1.43	1.41	1.39	1.36	1.32	1.30	1.23	1.19	1.15	300
500	1.56	1.54	1.52	1.50	1.48	1.45	1.42	1.40	1.38	1.34	1.30	1.28	1.21	1.16	1.11	500
1000	1.55	1.53	1.51	1.49	1.47	1.44	1.41	1.38	1.36	1.33	1.29	1.26	1.19	1.13	1.08	1000
∞	1.54	1.52	1.50	1.48	1.46	1.42	1.39	1.37	1.35	1.32	1.27	1.24	1.17	1.11	1.00	∞

$\alpha = 0.01$

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	n_1 / n_2
1	405	500	540	563	576	586	593	598	602	606	611	614	617	619	621	1
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	2
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.8	26.8	26.7	3
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.2	14.1	14.0	4
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.77	9.68	9.61	9.55	5
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.60	7.52	7.45	7.40	6
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.36	6.27	6.21	6.16	7
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.56	5.48	5.41	5.36	8
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	5.00	4.92	4.86	4.81	9
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.60	4.52	4.46	4.41	10
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.29	4.21	4.15	4.10	11
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.05	3.97	3.91	3.86	12
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.86	3.78	3.71	3.66	13
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.70	3.62	3.56	3.51	14
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.56	3.49	3.42	3.37	15
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.45	3.37	3.31	3.26	16
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.35	3.27	3.21	3.16	17
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.27	3.19	3.13	3.08	18
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.19	3.12	3.05	3.00	19
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.13	3.05	2.99	2.94	20
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.07	2.99	2.93	2.88	21
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	3.02	2.94	2.88	2.83	22
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.97	2.89	2.83	2.78	23
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.93	2.85	2.79	2.74	24
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.89	2.81	2.75	2.70	25

26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.86	2.78	2.72	2.66	26
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.82	2.75	2.68	2.63	27
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.79	2.72	2.65	2.60	28
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.77	2.69	2.63	2.57	29
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.74	2.66	2.60	2.55	30
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02	2.93	2.80	2.70	2.62	2.55	2.50	32
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98	2.89	2.76	2.66	2.58	2.51	2.46	34
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95	2.86	2.72	2.62	2.54	2.48	2.43	36
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92	2.83	2.69	2.59	2.51	2.45	2.40	38
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.56	2.48	2.42	2.37	40
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86	2.78	2.64	2.54	2.46	2.40	2.34	42
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84	2.75	2.62	2.52	2.44	2.37	2.32	44
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82	2.73	2.60	2.50	2.42	2.35	2.30	46
48	7.20	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80	2.72	2.58	2.48	2.40	2.33	2.28	48
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.56	2.46	2.38	2.32	2.27	50
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.39	2.31	2.25	2.20	60
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.42	2.31	2.23	2.17	2.12	80
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.37	2.26	2.19	2.12	2.07	100
125	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.66	2.55	2.47	2.33	2.23	2.15	2.08	2.03	125
150	6.81	4.75	3.92	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.31	2.20	2.12	2.06	2.00	150
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.27	2.17	2.09	2.02	1.97	200
300	6.72	4.68	3.85	3.38	3.08	2.86	2.70	2.57	2.47	2.38	2.24	2.14	2.06	1.99	1.94	300
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.22	2.12	2.04	1.97	1.92	500
1000	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.20	2.10	2.02	1.95	1.90	1000
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.08	2.00	1.93	1.88	∞

$\alpha = 0.01$

n_2 \ n_1	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞	n_1 / n_2
1	622	623	624	625	626	628	629	630	630	631	633	633	635	636	637	1
2	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.6	99.5	99.5	2
3	26.6	26.6	26.6	26.5	26.5	26.5	26.4	26.4	26.4	26.3	26.3	26.2	26.2	26.1	26.1	3
4	14.0	13.9	13.9	13.9	13.8	13.8	13.7	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	4
5	9.51	9.47	9.43	9.40	9.38	9.33	9.29	9.26	9.24	9.20	9.16	9.13	9.08	9.04	9.02	5
6	7.35	7.31	7.28	7.25	7.23	7.18	7.14	7.11	7.09	7.06	7.01	6.99	6.93	6.90	6.88	6
7	6.11	6.07	6.04	6.02	5.99	5.94	5.91	5.88	5.86	5.82	5.78	5.75	5.70	5.67	5.65	7
8	5.32	5.28	5.25	5.22	5.20	5.15	5.12	5.00	5.07	5.03	4.99	4.96	4.91	4.88	4.86	8
9	4.77	4.73	4.70	4.67	4.65	4.60	4.57	4.54	4.52	4.48	4.44	4.42	4.36	4.33	4.31	9
10	4.36	4.33	4.30	4.27	4.25	4.20	4.17	4.14	4.12	4.08	4.04	4.01	3.96	3.93	3.91	10
11	4.06	4.02	3.99	3.96	3.94	3.89	3.86	3.83	3.81	3.78	3.73	3.71	3.66	3.62	3.60	11
12	3.82	3.78	3.75	3.72	3.70	3.65	3.62	3.59	3.57	3.54	3.49	3.47	3.41	3.38	3.36	12
13	3.62	3.59	3.56	3.53	3.51	3.46	3.43	3.40	3.38	3.34	3.30	3.27	3.22	3.19	3.17	13
14	3.46	3.43	3.40	3.37	3.35	2.30	3.27	3.24	3.22	3.18	3.14	3.11	3.06	3.05	3.00	14
15	3.33	3.29	3.26	3.24	3.21	3.17	3.13	3.10	3.08	3.05	3.00	2.98	2.92	2.89	2.87	15
16	3.22	3.18	3.15	3.12	3.10	3.05	3.02	2.99	2.97	2.93	2.89	2.86	2.81	2.78	2.75	16
17	3.12	3.08	3.05	3.03	3.00	2.96	2.92	2.89	2.87	2.83	2.79	2.76	2.71	2.68	2.65	17
18	3.03	3.00	2.97	2.94	2.92	2.87	2.84	2.81	2.78	2.75	2.70	2.68	2.62	2.59	2.57	18
19	2.96	2.92	2.89	2.87	2.84	2.80	2.76	2.73	2.71	2.67	2.63	2.60	2.55	2.51	2.49	19
20	2.90	2.86	2.83	2.80	2.78	2.73	2.69	2.67	2.64	2.61	2.56	2.54	2.48	2.44	2.42	20
21	2.84	2.80	2.77	2.74	2.72	2.67	2.64	2.61	2.58	2.55	2.50	2.48	2.42	2.38	2.36	21
22	2.78	2.75	2.72	2.69	2.67	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.36	2.33	2.31	22
23	2.74	2.70	2.67	2.64	2.62	2.57	2.54	2.51	2.48	2.45	2.40	2.37	2.32	2.28	2.26	23
24	2.70	2.66	2.63	2.60	2.58	2.53	2.49	2.46	2.44	2.40	2.36	2.33	2.27	2.24	2.21	24
25	2.66	2.62	2.59	2.56	2.54	2.49	2.45	2.42	2.40	2.36	2.32	2.29	2.23	2.19	2.17	25

26	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.39	2.36	2.33	2.28	2.25	2.19	2.16	2.13	26
27	2.59	2.55	2.52	2.49	2.47	2.42	2.38	2.35	2.33	2.29	2.25	2.22	2.16	2.12	2.10	27
28	2.56	2.52	2.49	2.46	2.44	2.39	2.35	2.32	2.30	2.26	2.22	2.19	2.13	2.09	2.06	28
29	2.53	2.49	2.46	2.44	2.41	2.36	2.33	2.30	2.27	2.23	2.19	2.16	2.10	2.06	2.03	29
30	2.51	2.47	2.44	2.41	2.39	2.34	2.30	2.27	2.25	2.21	2.16	2.13	2.07	2.03	2.01	30
32	2.46	2.42	2.39	2.36	2.34	2.29	2.25	2.22	2.20	2.16	2.11	2.08	2.02	1.98	1.96	32
34	2.42	2.38	2.35	2.32	2.30	2.25	2.21	2.18	2.16	2.12	2.07	2.04	1.98	1.94	1.91	34
36	2.38	2.35	2.32	2.29	2.26	2.21	2.17	2.14	2.12	2.08	2.03	2.00	1.94	1.90	1.87	36
38	2.35	2.32	2.28	2.26	2.23	2.18	2.14	2.11	2.09	2.05	2.00	1.97	1.90	1.86	1.84	38
40	2.33	2.29	2.26	2.23	2.20	2.15	2.11	2.08	2.06	2.02	1.97	1.94	1.87	1.83	1.80	40
42	2.30	2.26	2.23	2.20	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.91	1.85	1.80	1.78	42
44	2.28	2.24	2.21	2.18	2.15	2.10	2.06	2.03	2.01	1.97	1.92	1.89	1.82	1.78	1.75	44
46	2.26	2.22	2.19	2.16	2.13	2.08	2.04	2.01	1.99	1.95	1.90	1.86	1.80	1.76	1.73	46
48	2.24	2.20	2.17	2.14	2.12	2.06	2.02	1.99	1.97	1.93	1.88	1.84	1.78	1.73	1.70	48
50	2.22	2.18	2.15	2.12	2.10	2.05	2.01	1.97	1.95	1.91	1.86	1.82	1.76	1.71	1.68	50
60	2.15	2.12	2.08	2.05	2.03	1.98	1.94	1.90	1.88	1.84	1.78	1.75	1.68	1.63	1.60	60
80	2.07	2.03	2.00	1.97	1.94	1.89	1.85	1.81	1.79	1.75	1.69	1.66	1.58	1.53	1.49	80
100	2.02	1.98	1.94	1.92	1.89	1.84	1.80	1.76	1.73	1.69	1.63	1.60	1.52	1.47	1.43	100
125	1.98	1.94	1.91	1.88	1.85	1.80	1.76	1.72	1.69	1.65	1.59	1.55	1.47	1.41	1.37	125
150	1.96	1.92	1.88	1.85	1.83	1.77	1.73	1.69	1.66	1.62	1.56	1.52	1.43	1.38	1.33	150
200	1.93	1.89	1.85	1.82	1.79	1.74	1.69	1.66	1.63	1.58	1.52	1.48	1.39	1.33	1.28	200
300	1.89	1.85	1.82	1.79	1.76	1.71	1.66	1.62	1.59	1.55	1.48	1.44	1.35	1.28	1.22	300
500	1.87	1.83	1.79	1.76	1.74	1.68	1.63	1.60	1.56	1.52	1.45	1.41	1.31	1.23	1.16	500
1000	1.85	1.81	1.77	1.74	1.72	1.66	1.61	1.57	1.54	1.50	1.43	1.38	1.28	1.19	1.11	1000
∞	1.83	1.79	1.76	1.72	1.70	1.64	1.59	1.55	1.52	1.47	1.40	1.36	1.25	1.15	1.00	∞

表 4. 检验相关系数 $\rho=0$ 的临界值(r_α)表

$$P(|r|>r_\alpha)=\alpha$$

n \ α	0.10	0.05	0.02	0.01	0.001	α / n
1	0.98769	0.99692	0.999507	0.999877	0.9999988	1
2	.90000	.95000	.98000	.99000	.99900	2
3	.8054	.8783	.93433	.95873	.99116	3
4	.7293	.8114	.8822	.91720	.97406	4
5	.6694	.7545	.8329	.8745	.95074	5
6	.6215	.7067	.7887	.8343	.92493	6
7	.5822	.6664	.7498	.7977	.8982	7
8	.5494	.6319	.7155	.7646	.8721	8
9	.5214	.6021	.6851	.7348	.8471	9
10	.4973	.5760	.6581	.7079	.8233	10
11	.4762	.5529	.6339	.6835	.8010	11
12	.4575	.5324	.6120	.6614	.7800	12
13	.4409	.5139	.5923	.6411	.7603	13
14	.4259	.4973	.5742	.6226	.7420	14
15	.4124	.4821	.5577	.6055	.7246	15
16	.4000	.4683	.5425	.5897	.7084	16
17	.3887	.4555	.5285	.5751	.6932	17
18	.3783	.4438	.5155	.5614	.6787	18
19	.3687	.4329	.5034	.5487	.6652	19
20	.3598	.4227	.4921	.5368	.6524	20
25	.3233	.3809	.4451	.4869	.5974	25
30	.2960	.3494	.4093	.4487	.5541	30
35	.2746	.3246	.3810	.4182	.5189	35
40	.2573	.3044	.3578	.3932	.4896	40
45	.2428	.2875	.3384	.3721	.4648	45
50	.2306	.2732	.3218	.3541	.4433	50
60	.2108	.2500	.2948	.3248	.4078	60
70	.1954	.2319	.2737	.3017	.3799	70
80	.1829	.2172	.2565	.2830	.3568	80
90	.1726	.2050	.2422	.2673	.3375	90
100	.1638	.1946	.2301	.2540	.3211	100

表5. 随机数表

03	47	43	73	86	36	96	47	36	61	46	98	63	71	62	33	26	16	80	45	60	11	14	10	95
97	74	24	67	62	42	81	14	57	20	42	53	32	37	32	27	07	36	07	51	24	51	79	89	73
16	76	62	27	66	56	50	26	71	07	32	90	79	78	53	13	55	38	58	59	88	97	54	14	10
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15	57	12	10	14	21	88	26	49	81	76
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90	06	18	44	32	53	23	83	01	30	30
16	22	77	94	39	49	54	43	54	82	17	37	93	23	78	87	35	20	96	43	84	26	34	91	64
84	42	17	53	31	57	24	55	06	88	77	04	74	47	67	21	76	33	50	25	83	92	12	06	76
63	01	63	78	59	16	95	55	67	19	98	10	50	71	75	12	86	73	58	07	44	39	52	38	79
33	21	12	34	29	78	64	56	07	82	52	42	07	44	38	15	51	00	13	42	99	66	02	79	54
57	60	86	32	44	09	47	27	96	54	49	17	46	09	62	90	52	84	77	27	08	02	73	43	28
18	18	07	92	45	44	17	16	58	09	79	83	86	19	62	06	76	50	03	10	55	23	64	05	05
26	62	38	97	75	84	16	07	44	99	83	11	46	32	24	20	14	85	88	45	10	93	72	88	71
23	42	40	64	74	82	97	77	77	81	07	45	32	14	08	32	98	94	07	72	93	85	79	10	75
52	36	28	19	95	50	92	26	11	97	00	56	76	31	38	80	22	02	53	53	86	60	42	04	53
37	85	94	35	12	83	39	50	08	30	42	34	07	96	88	54	42	06	87	98	35	85	29	48	39
70	29	17	12	13	40	33	20	38	26	13	89	51	03	74	17	76	37	13	04	07	74	21	19	30
56	62	18	37	35	96	83	50	87	75	97	12	25	93	47	70	33	24	03	54	97	77	46	44	80
99	49	57	22	77	88	42	95	45	72	16	64	36	16	00	04	43	18	66	79	94	77	24	21	90
16	08	15	04	72	33	27	14	34	09	45	59	34	68	49	12	72	07	34	45	99	27	72	95	14
31	16	93	32	43	50	27	89	87	19	20	15	37	00	49	52	85	66	60	44	38	68	88	11	80
68	34	30	13	70	55	74	30	77	40	44	22	78	84	26	04	33	46	09	52	68	07	97	06	57
74	57	25	65	76	59	29	97	68	60	71	91	38	67	54	13	58	18	24	76	15	54	55	95	52
27	42	37	86	53	48	55	90	65	72	96	57	69	36	10	96	46	92	42	45	97	60	49	04	91
00	39	68	29	61	66	37	32	20	30	77	84	57	03	29	10	45	65	04	26	11	04	96	67	24
29	94	98	94	24	68	49	69	10	82	53	75	91	93	30	34	25	20	57	27	40	48	73	51	92

16	90	82	66	59	83	62	64	11	12	67	19	00	71	74	60	47	21	19	68	02	02	37	03	31
11	27	94	75	06	06	09	19	74	66	02	94	37	34	02	76	70	90	30	86	38	45	94	30	38
35	24	10	16	20	33	32	51	26	38	79	78	45	04	91	16	92	53	56	16	02	75	50	95	98
38	23	16	86	38	42	38	97	01	50	87	75	66	81	41	40	01	74	91	62	48	51	84	08	32
31	96	25	91	47	96	44	33	49	13	34	86	82	53	91	00	52	43	48	85	27	55	26	89	62
66	67	40	67	14	64	05	71	95	86	11	05	65	09	68	76	83	20	37	90	57	16	00	11	66
14	90	84	45	11	75	73	88	05	90	52	27	41	14	86	22	98	12	22	08	07	52	74	95	80
68	05	51	18	00	33	96	02	75	19	07	60	62	93	55	59	33	82	43	90	49	37	38	44	59
20	46	78	73	90	97	51	40	14	02	04	02	33	31	08	39	54	16	49	36	47	95	93	13	30
64	19	58	97	79	15	06	15	93	20	01	90	10	75	06	40	78	78	89	62	02	67	74	17	33
05	26	93	70	60	22	35	85	15	13	92	03	51	59	77	59	56	78	06	83	52	91	05	70	74
07	97	10	88	23	09	98	42	99	64	61	71	62	99	15	06	51	29	16	93	58	05	77	09	51
68	71	86	85	85	54	87	66	47	54	73	32	08	11	12	44	95	92	63	16	29	56	24	29	48
26	99	61	65	53	58	37	78	80	70	42	10	50	67	42	32	17	55	86	74	94	44	67	16	94
14	65	52	68	75	87	59	36	22	41	26	78	63	06	55	13	08	27	01	50	15	29	39	39	43
17	53	77	58	71	71	41	61	50	72	12	41	94	96	26	44	96	27	36	99	02	96	74	30	83
90	26	59	21	19	23	52	23	33	12	96	93	02	18	39	07	02	18	36	07	25	99	32	70	23
41	23	52	55	99	31	04	49	69	96	10	47	48	45	88	13	41	43	89	20	97	17	14	49	17
60	20	50	81	69	31	99	73	68	68	35	81	33	03	76	24	30	12	48	60	18	99	10	72	34
91	25	38	05	90	94	58	28	41	36	45	37	59	03	09	90	35	57	29	12	82	62	54	65	60
34	50	57	74	37	98	80	33	00	91	09	77	93	19	82	74	94	80	04	04	45	07	31	66	49
85	22	04	39	43	73	81	53	94	79	33	62	46	86	28	08	31	54	46	31	53	94	13	38	47
09	79	13	77	48	73	82	97	22	21	05	03	27	24	83	72	89	44	05	60	35	80	39	94	88
88	75	80	18	14	22	95	75	42	49	39	32	82	22	49	02	48	07	70	37	16	04	61	67	87
90	96	23	70	60	39	00	03	06	90	55	85	78	38	36	94	37	30	69	32	90	89	00	76	33

53	74	23	99	67	61	32	28	69	84	94	62	67	86	24	98	33	41	19	95	47	53	53	38	09
63	38	06	86	54	99	00	65	26	94	02	82	90	23	07	79	62	67	80	60	75	91	12	81	19
35	30	58	21	46	06	72	17	10	94	25	21	31	75	96	49	28	24	00	49	55	65	79	78	07
63	43	36	82	69	65	51	18	37	88	61	38	44	12	45	32	92	85	88	65	54	34	81	85	35
98	25	37	55	26	01	91	82	81	46	74	71	12	94	97	24	02	71	37	07	03	92	18	66	75
02	63	21	17	69	71	50	80	89	56	38	15	70	11	48	43	40	45	86	98	00	83	26	91	03
64	55	22	21	82	48	22	28	06	00	61	54	13	43	91	82	78	12	23	29	06	66	24	12	27
85	07	26	13	39	01	10	07	82	04	59	63	69	36	03	69	11	15	83	80	13	29	54	19	28
58	54	16	24	15	51	54	44	82	00	62	61	65	04	69	38	18	65	18	97	85	72	13	49	21
34	85	27	84	87	61	48	64	56	26	90	18	48	13	26	37	70	15	42	57	65	65	80	39	07
03	92	18	27	46	57	99	16	96	56	30	33	72	85	22	84	64	38	56	98	99	01	30	98	64
62	93	30	27	59	37	75	41	66	48	86	97	80	61	45	23	53	04	01	63	45	76	08	64	27
08	45	93	15	22	60	21	75	46	91	98	77	27	85	42	28	88	61	08	84	69	62	03	42	73
07	08	55	18	40	45	44	75	13	90	24	94	96	61	02	57	55	66	83	15	73	42	37	11	61
01	85	89	95	66	51	10	19	34	88	15	84	97	19	75	12	76	39	43	78	64	63	91	08	25
72	84	71	14	35	19	11	58	49	26	50	11	17	17	76	86	31	57	20	18	95	60	78	46	75
88	78	28	16	84	13	52	53	94	53	75	45	69	30	96	73	89	65	70	31	99	17	43	48	76
45	17	75	65	57	28	40	19	72	12	25	12	74	75	67	60	40	60	81	19	24	62	01	61	16
96	76	28	12	54	22	01	11	94	25	71	96	16	16	88	68	64	36	74	45	19	59	50	88	92
43	31	67	72	30	24	02	94	08	63	38	32	36	66	02	69	36	38	25	39	48	03	45	15	22
50	44	66	44	21	66	06	58	05	62	68	15	54	35	02	42	35	48	96	32	14	52	41	52	48
22	66	22	15	86	26	63	75	41	99	58	42	36	72	24	58	37	52	18	51	03	37	18	39	11
96	24	40	14	51	23	22	30	88	67	95	67	47	29	83	94	69	40	06	07	18	16	36	78	86
31	73	91	61	19	60	20	72	93	48	98	57	07	23	69	65	95	59	69	58	56	80	30	19	44
78	60	73	99	84	43	89	94	36	45	56	69	47	07	41	90	22	91	07	12	78	35	34	08	72

84	37	90	61	56	70	10	23	98	05	85	11	34	76	60	76	48	45	34	60	01	64	18	39	96
36	67	10	08	23	98	93	35	08	86	99	29	76	29	81	33	34	91	58	98	63	14	52	32	52
07	28	59	07	48	89	64	58	89	75	83	85	62	27	89	80	14	78	56	27	86	63	59	80	02
10	15	83	87	60	79	24	31	66	56	21	48	24	06	93	91	98	94	05	49	01	47	59	38	00
55	19	68	97	65	03	73	52	16	56	00	53	55	90	27	33	42	29	38	87	22	13	88	83	34
53	81	29	13	39	35	01	20	71	34	62	33	74	82	14	53	73	19	09	03	56	54	29	56	93
51	86	32	68	92	33	98	74	66	99	40	14	71	94	58	45	94	19	38	81	14	44	99	81	07
35	91	70	29	13	80	03	54	07	27	96	94	78	32	66	50	95	52	74	33	13	80	55	62	54
37	71	67	95	13	20	02	44	95	94	64	85	04	05	72	01	32	90	76	14	53	89	74	60	41
93	66	13	83	27	92	79	64	64	72	28	54	96	53	84	48	14	52	98	94	56	07	93	89	30
02	96	08	45	65	13	05	00	41	84	93	07	54	72	59	21	45	57	09	77	19	48	56	27	44
49	83	43	48	35	82	88	33	69	96	72	36	04	19	76	47	45	15	18	60	82	11	08	95	97
84	60	71	62	46	40	80	81	30	37	34	39	23	05	38	25	15	35	71	30	88	12	57	21	77
18	17	30	88	71	44	91	14	88	47	89	23	30	63	15	56	34	20	47	89	99	82	93	24	98
79	69	10	61	78	71	32	76	95	62	87	00	22	58	40	92	54	01	75	25	43	11	71	99	31
75	93	36	57	83	56	20	14	82	11	74	21	97	90	65	96	42	68	63	86	74	54	13	26	94
38	30	92	29	03	06	28	81	39	38	62	25	06	84	63	61	29	08	93	67	04	32	92	08	09
51	29	50	10	34	31	57	75	95	80	51	97	02	74	77	76	15	48	49	44	18	55	63	77	09
21	31	38	86	24	37	79	81	53	74	73	24	16	10	33	52	83	90	94	76	70	47	14	54	36
29	01	23	87	88	58	02	39	37	67	42	10	14	20	92	16	55	23	42	45	54	96	09	11	06
95	33	95	22	00	18	74	72	00	18	38	79	58	69	32	81	76	80	26	92	82	80	84	25	39
90	84	60	79	80	24	36	59	87	38	82	07	53	89	35	96	35	23	79	18	05	98	90	07	35
46	40	62	98	82	54	97	20	56	95	15	74	80	08	32	16	46	70	50	80	67	72	16	42	79
20	31	89	03	43	38	46	82	68	72	32	14	82	99	70	80	60	47	18	97	63	49	30	21	30
71	59	73	05	50	08	22	23	71	77	91	01	93	20	49	82	96	59	26	94	66	39	67	98	60

译名对照表[①]

Brown	布朗
Cochran	柯赫伦
David	达威
Dugue	杜琪
Fisher	费歇
Geiger	盖革
Monte-Carlo	蒙特卡罗
Neymann	奈曼
Rao	罗
Scheffe	许切非
Wald	瓦尔德
Wiener	维纳
Гливенко	格里汶科

① 凡在上册对照表中出现过的外国人名，下册都未列入。

参考书目

[1] 概率论基础及其应用　王梓坤，科学出版社(1976)

[2] 概率论与数理统计　复旦大学数学系，上海科学技术出版社(1961)

[3] 基础概率与数理统计　林少宫，人民教育出版社(1963)

[4] 工业技术应用数理统计学　周华章，高等教育出版社(上册 1960，下册 1964)

[5] 概率论及数理统计　M. Fisz，王福保译，上海科学技术出版社(1962)

[6] 统计数学方法　H. Cramer，魏宗舒等译，上海科学技术出版社(1966)

[7] 统计分析　森口繁一(1957)刘璋温译

[8] 概率论　伊藤清(1952)，刘璋温译

[9] 平稳随机序列统计分析

N. Grenander, M. Rosenblatt (1957)，郑绍濂等译，上海科学技术出版社(1962)

[10] 平稳随机函数导论

A. M. Яглом 数学进展(1956)，梁之舜译

[11] Probability, Random Variables and Stochastic Processes

A. Papoulis, McGraw-Hill books Co. (1965)，有中译本，谢国瑞等译，高等教育出版社(1983)

[12] Introduction to Probability and Statistics

Part II: Statistics　N. G. Giri, Marcel Dekker Inc., New York(1975)

[13] Stochastic Processes　E. Parzen(1967)

有中译本，邓永录，杨振明译，高等教育出版社(1988)

[14] Mathematical Statistics　S. S. Wilks(1962)

[15] Testing statistics Hypotheses　E. L. Lehmann(1959)

[16] 常用数理统计方法　中国科学院数学研究所　统计组，科学出版社(1973)

[17] 回归分析方法　中国科学院数学研究所　统计组，科学出版社(1975)

[18] 数理统计引论　　陈希孺，科学出版社(1981)

[19] 参数估计　　成平等译，上海科学技术出版社(1985)

[20] 试验的分析与设计　　H.B.Mann，张里千等译，科学出版社(1963)

[21] Linear Statistics Inference and Its Applications
Second Edition　　C.R.Rao(1973)

[22] 随机过程论　　王梓坤，科学出版社(1965)

[23] Random Differential Equation in Science and Engineering
T.T.Soong(1973)

[24] Random Point Processes　　D.L.Snyder(1975)，有中译本，梁之舜、邓永录译，高等教育出版社(1981)

[25] 公用事业理论的数学方法
А.Я.Хинчин(1955)，张里千等译，科学出版社(1960)

[26] Markov Chains with Stationary Transition Probabilities
K.L.Chung，Springer-Verlag，2nd ed(1967)

[27] 回归分析及其试验设计
上海师范大学概率统计教研组，上海教育出版社(1978)

[28] 概率论　　复旦大学信息论教研室(1979)

[29] 统计学导论
A.M.Mood & F.A.Graybill(1963)，史定华译，科学出版社(1978)

[30] 广义函数的基本理论　　J.Mikusinski R.Sikorski
(1960)，复旦大学数学系译，人民教育出版社(1960)

[31] 测度与概率基础　　中山大学《测度与概率基础编写组》，
广东科学技术出版社(1981)

[32] 随机过程(I)　　[苏]И.И.基赫曼，A.B.斯科罗霍得，
邓永录　邓集贤　石北源译，科学出版社(1985)

[33] 随机过程(III)　　[苏]И.И.基赫曼，A.B.斯科罗霍得，
石北源等译，科学出版社(即将出版)

[34] 概率统计计算　　中国科学院计算中心概率统计组，科学出版社(1979)

[35] 军事技术运筹学基础　　[苏]Ю.B.楚耶夫等著，冷拓等译，
国防工业出版社(1976)

下册习题答案

第　六　章

1. (1) 0.132; (2) 0.582; (3) 0.291.

2. $P\{M=m\}=(1-q^{m+1})^n-(1-q^m)^n \quad m=0,1,2,\cdots.$

   $P\{K=k\}=pq^{nk} \quad k=0,1,2,\cdots.$

3. (1) $e^{-7.2}$; (2) $(1-e^{-4.5})^5$.

5. 提示：参考第三章的习题 25. 记 $\eta_i=\dfrac{\xi_i-a}{\sigma}$，则 η_i 服从正态

$N(0,1)$, $i=1,2,\cdots,n$.

6. 0.671.

9. 提示：记 $\zeta_i=\xi_i-\eta_i$, $i=1,2,\cdots,n$. 对 $\zeta_1,\cdots,\zeta_n$ 应用定理 6.3.1.

13. $\eta\sim t_{(n-1)}$.

7. $$f_D(y)=\begin{cases}\dfrac{1}{\sqrt{\pi}\,\sigma}e^{-\frac{y^2}{4\sigma^2}}, & y\geqslant 0;\\ 0, & y<0,\end{cases}$$

$$f_n(u)=\frac{1}{\pi\sigma}e^{-\frac{(u-a)^2}{2\sigma^2}}\int_0^{\frac{u-a}{\sigma}}e^{-\frac{x^2}{2}}dx,$$

$$f_1(v)=\frac{2}{\sqrt{2\pi}\,\sigma}e^{-\frac{(v-a)^2}{2\sigma^2}}-f_n(u).$$

8. $f_n(u)=2\lambda e^{-\lambda u}[1-e^{-\lambda u}], u>0;$

   $f_1(v)=2\lambda e^{-2\lambda v}, \qquad v>0;$

   $f_D(y)=\lambda e^{-\lambda y}, \qquad y>0.$

16. $f_n(u)=n\left[u-\left(\theta-\dfrac{1}{2}\right)\right]^{n-1}$;

    $f_1(v)=n\left[\left(\theta+\dfrac{1}{2}\right)-v\right]^{n-1}$;

    $f(u,v)=n(n-1)(u-v)^{n-1}$.

第 七 章

1. 提示：$P\{\bar{\xi}\geqslant M\}=0.01$，$2n\bar{\xi}/\beta(\alpha+1)\sim\chi^2_{2n}$.

2. $1-e^{-\beta M}, M\geqslant 0$.

3. -0.525.

4. (1) 当 $n=2k+1$ 时 $\hat{\theta}=\xi^*_{k+1}$；当 $n=2k$ 时，$\xi^*_k<\hat{\theta}<\xi^*_{k+1}$；

   (2) $\hat{\theta}=\dfrac{-1}{n\sum\limits_{i=1}^{n}\ln\xi_i}$； (3) $\hat{\theta}=\max\limits_{1\leqslant i\leqslant n}\xi_i$； (4) $\hat{\theta}=\bar{\xi}$.

5. $\hat{p}=1-\dfrac{S^2}{\bar{\xi}}$，$\hat{N}=\bar{\xi}/\hat{p}$.

6. $\hat{a}=\lg\bar{\xi}-\dfrac{\ln 10}{2}\hat{\sigma}^2$，$\hat{\sigma}^2=\dfrac{1}{\ln 10}\ln\dfrac{S^2-\bar{\xi}^2}{\bar{\xi}^2}$.

7. 1.834.

8. (1) $\hat{\beta}=\dfrac{1}{\bar{T}-\hat{t}_0}$；(2) $\hat{t}_0=\min\limits_{1\leqslant i\leqslant n}T_i$.

9. $\hat{\lambda}=n\Big/\sum\limits_{i=1}^{n}\xi_i^{\alpha}$.

10. $\hat{\beta}=\bar{\xi}/\alpha+1$.

11. $\hat{\theta}_1=\min\limits_{1\leqslant i\leqslant n}\xi_i$，$\hat{\theta}_2=\bar{\xi}-\min\limits_{1\leqslant i\leqslant n}\xi_i$.

12. (1) $k=n\sqrt{\dfrac{2}{\pi}}$；(2) $k=2(n-1)$.

13. 提示 $\bar{\xi}=\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i$， $T=\sum\limits_{i=1}^{n}\alpha_i\xi_i$， $\sum\limits_{i=1}^{n}\alpha_i=1$.

    可证 $E(\bar{\xi}T)=\dfrac{\sigma^2}{n}+a^2$.

14. 提示：利用(7.2.16)式，求得 σ 的优效估计量的方差为$\dfrac{\sigma^2}{2n}$.

15. $\hat{a}_2$ 的方差最小.

16. $c_1=(\sigma_1^2-\rho\sigma_1\sigma_2)/(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)$；

    $c_2=(\sigma_2^2-\rho\sigma_1\sigma_2)/(\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2)$.

17. $\hat{a}=\bar{\xi}-\bar{\eta}$，$n_1=\left[\dfrac{n}{3}\right]$ ([ ]表示取整数部分).

19. (1) $\max\limits_{1\leq i\leq n}\xi_i-\frac{1}{2}\leq\theta\leq\min\limits_{1\leq i\leq n}\xi_i+\frac{1}{2}$;

(2) $\frac{1}{2}[\max\limits_{1\leq i\leq n}\xi_i+\min\limits_{1\leq i\leq n}\xi_i]$较$\xi$有效

20. $\frac{4}{3}\max\limits_{1\leq i\leq 3}\xi_i$ 较 $4\min\limits_{1\leq i\leq 3}\xi_i$ 有效。

21. $k_1=\frac{1}{3}$, $k_2=\frac{2}{3}$.

25. (1) $\frac{n_1S_1^2\sigma_2^2(n_2-1)}{n_2S_2^2\sigma_1^2(n_1-1)}\sim F_{(n_1-1,n_2-1)}$.

(2) $\frac{(\bar{\xi}-\bar{\eta})-(a_1-a_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ 服从正态 $N(0,1)$;

26. $P\left\{(\bar{\xi}-\bar{\eta})-t_{(n-1)}(0.05)\frac{S}{\sqrt{n-1}}<a_1-a_2\right.$

$\left.<(\bar{\xi}-\bar{\eta})+t_{(n-1)}(0.05)\frac{S}{\sqrt{n-1}}\right\}=1-\alpha$,

其中 $S^2=S_1^2+S_2^2+2S_{12}$.

27. $(-0.07, 1.23)$; $(0.726, 3.63)$.

28. $\bar{x}=2$, $S^*=\frac{\sqrt{52}}{3}$; $(0.607, 3.393)$, $(3.415, 17.707)$.

29. $\lambda=\frac{n+\alpha}{T+\frac{1}{\beta}}$, $T=\sum\limits_{i=1}^{n}\xi_i$.

30. $R(a, d_1)=\frac{39}{25}a^2$, $R(a, d_2)=\frac{13}{6}a^2$, $R(a, d_3)=\frac{5}{3}a^2$.

第 八 章

1. 算得 $t_{(10+12-2)}=-0.065$,查表得 $t_{20}(0.05)=2.09$, 由于$|-0.065|<2.09$,所以可认为平均值之间的差异不显著.

2. 算得 $F_{(5,8)}=1.03$,查表得 $F_{(5,8)}(0.05)=3.69$,由于 $1.03<3.69$,所以可认为方差之间无显著性的差异.

3. 算得 $\chi^2_{8-1-1}=1.254$,查表得 $\chi^2_6(0.05)=12.59$,由于 $1.254<12.59$,所以可认为服从泊松分布.

4. 算得 $\chi^2_{6-1-1}=1.551$,查表得 $\chi^2_4(0.10)=9.236$,由于 $1.551<9.236$,

所以可认为服从二项分布.

5. 算得 $\chi^2_{10-1}=5.125$,查表得 $\chi^2_9(0.10)=15.987$,由于 $5.125<15.987$,所以可认为服从均匀分布.

6. 算得 $\chi^2_{11-1-1}=1.1929$,查表得 $\chi^2_9(0.10)=15.987$,由于 $1.1929<15.987$,所以可认为服从二项分布.

10. 当原假设 H_0 成立时,2ζ 服从正态 $N(0,1)$.

11. 记 $\zeta_n=\sum_{i=1}^{n}\xi_i$,当原假设 H_0 成立时,$\zeta_n\sim\chi^2_n$,

(1) $P\{\zeta_n\geqslant(4\ln c+4n\ln 2)\}=0.05$,

(2) $P\{\zeta_n\leqslant(2n\ln 2-2\ln c)\}=0.05$.

给定显著性水平 $\alpha=0.05$ 及子样的容量 n,在 χ^2-分布临界值表中查得 $\chi^2_n(0.05)$ 值,由这个值可确定常数 c 值.

12. 记 $\zeta_n=\sum_{i=1}^{n}\xi_i$,则 ζ_n 服从二项分布 $b(n;p)$,其中 $p=\theta$. 则

(1) 当 H_0 成立时,ζ_n 服从 $b\left(n;\frac{1}{2}\right)$,常数 c 由下式确定:

$$P\left\{\zeta_n\leqslant\frac{1}{\ln 2}\left[n\ln\frac{4}{3}-\ln c\right]\right\}=0.05.$$

(2) 当 H_0 成立时,ζ_n 服从 $b\left(n;\frac{1}{3}\right)$,常数 c 由下式确定:

$$P\left\{\zeta_n\geqslant\frac{1}{\ln 2}\left[n\ln\frac{4}{3}+\ln c\right]\right\}=0.05.$$

13. 记 $\zeta_n=\sum_{i=1}^{n}\xi_i$,则 ζ_n 服从参数为 $n\lambda$ 的泊松分布. 当 H_0 成立时,ζ_n 的概率函数为:

$$P\{\zeta_n=x\}=\frac{(\theta n)^x}{x!}e^{-\theta n},$$

给定显著性水平 $\alpha=0.05$,常数 c 由下式确定:

$$P\left\{\zeta_n\geqslant\frac{1}{\ln 10}[0.9n+\ln c]\right\}=0.05.$$

14. $n=250$.

15. $n=11$.

第 九 章

1. (1) $\hat{\beta}_0=-0.35$, $\hat{\beta}_1=0.81$.

(2) 在显著性水平 $\alpha=0.05$ 下，否定 H_0.

(3) $\beta_{11}=0.73$, $\beta_{12}=0.89$.

(4) $\hat{\sigma}_*^2=0.33$.

(5) $$\hat{y}_1=\hat{y}-0.2541\sqrt{1+\sum_{i=0}^{1}\sum_{j=0}^{1}c_{ij}x_ix_j},$$

$$\hat{y}_2=\hat{y}+0.2541\sqrt{1+\sum_{i=0}^{1}\sum_{j=0}^{1}c_{ij}x_ix_j},$$

其中 $\hat{y}=-0.35+0.81x$,

$$\sum_{i=0}^{1}\sum_{j=0}^{1}c_{ij}x_ix_j=\frac{91.51}{817}-\frac{19.8}{817}x+\frac{10}{817}x^2.$$

2. $\hat{y}=22.44-53.25\frac{1}{x_1}+0.34x_2$.

8. $\hat{\theta}=\frac{1}{5}(\xi_1+2\xi_2)$, $Q_e=\frac{1}{5}(2\xi_1-\xi_2)^2$.

9. $$\hat{\beta}=\begin{pmatrix}\frac{1}{3}(y_1+y_2+y_3)\\ -\frac{1}{2}(y_1-y_2)\\ \frac{1}{6}(y_1-2y_2+y_3)\end{pmatrix},$$

当 $\beta_2=0$ 时，则有

$$=\hat{\beta}\begin{pmatrix}\frac{1}{3}(y_1+y_2+y_3)\\ -\frac{1}{2}(y_1-y_2)\end{pmatrix}.$$

10. $\hat{\theta}=\bar{y}$, $\hat{\phi}=\frac{1}{3}[\bar{y}_{(2)}-\bar{y}_{(1)}]$,

其中 $\bar{y}=\frac{1}{5n}\sum_{i=1}^{5n}y_i$,

$$\bar{y}_{(2)}=\frac{1}{2n}\sum_{i=1}^{2n}y_{2n+i},$$

$$\bar{y}_{(3)}=\frac{1}{n}\sum_{i=1}^{n}y_{4n+i}.$$

11. $\hat{a}=\frac{1}{6}(y_1+2y_2+y_3),\hat{b}=\frac{1}{5}(-y_2+2y_3)$.

12. $\beta=-11, \beta_1=37$，否定 H_0.

13. $\beta_1=9.7, \beta_2=5.0, \beta_3=4.4, \beta_4=1.1$.